INRIA

CENTRE DE RENNES
IRISA

# Rapports de Recherche

N° 219

# BLIND EQUALIZERS

Albert BENVENISTE
Maurice GOURSAT

Juillet 1983

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(3) 954 90 20

IRIA

CENTRE DE RENNES
IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.: (3) 954 90 20

Rapports de Recherche

N° 219

# BLIND EQUALIZERS

Albert BENVENISTE
Maurice GOURSAT

Juillet 1983

# BLIND EQUALIZERS

Albert BENVENISTE
IRISA

Maurice GOURSAT
INRIA

## ABSTRACT

Blind equalizers do not require any known training sequence for the startup period but can rather perform at any time the equalization directly on the data stream. This report is divided in two parts. The first and main one presents a general approach for designing efficient blind equalizers for one and two carrier transmission systems. For simplicity this part is only concerned with transversal equalizers. Adjusting recursive equalizers is presented in the second part.

## RESUME

Les égaliseurs aveugles ne nécessitent pas de séquence d'apprentissage connue a priori mais peuvent s'accrocher à tout moment directement sur le flot de données. Ce rapport comporte deux parties distinctes. Dans la première, qui forme l'essentiel du rapport, nous présentons une approche générale pour construire des égaliseurs aveugles efficaces pour les systèmes de transmission à une ou deux porteuses. Pour simplifier la présentation nous n'avons considéré ici que le cas des égaliseurs transversaux. La méthode s'étend sans problème au cas des égaliseurs récursifs, ce qui est l'objet de la seconde partie.

# 1. BLIND EQUALIZERS

## I - INTRODUCTION.

Conventional equalization and carier recovery algorithms for minimizing mean-square error in digital communication systems generally require an initial training period during which a known data sequence is transmitted and synchronized at the receiver.

Equalizers for which such an initial training period can be avoided are referred to as blind equalizers. The need for blind equalizers in the field of data communication is widely discussed by D. Godard in (7), and we refer the reader to this paper for such a discussion.

The class of blind equalization algorithms we shall present in this paper have the following properties :

- In the case of a two carrier transmission system, they perform joint equalization and carrier recovery.

- The only change with respect to the classical algorithms lies in a modification of the standard error signal (i.e. difference between the output of the {equalizer + phase estimator} and the decoded data) which is used in such algorithms ; hence there is no increase of the computational complexity in comparison with the classical algorithms ; furthermore any classical structure may be used for the equalizer (transversal or recursive, with a stochastic gradient or a self-orthogonalizing type of adjustment).

- There is only a small increase of the duration of the blind startup period in comparison with classical startup periods involving a known training sequence.

- These blind equalizers are provided with a (smooth) automatic switching from the startup period to the standard transmission mode, and an automatic switching back to the blind startup period when an abrupt change occurs in the characteristics of the channel, without the need of any special testing procedure.

Hence the blind equalizers we present in this paper can be considered as extensions of classical adaptive equalizers.

Some modems which perform the blind equalization are presently available, but the litterature is poor with this problem. Let us mention the pioneering work of Sato (6) which has first designed such an equalizer for one carrier transmission systems, without any theoretical justification. D. Godard (7) has designed blind equalizers for two-carrier transmission systems, by decoupling the phase estimation and the removal of the intersymbol interference (ISI), thus resulting in a severe degradation with respect to the classical systems. The present authors have designed blind equalizers for one-carrier (8) (9) (13), and two-carrier (8), (11) transmission systems, without taking into account the problem of phase recovery. The present algorithms are based upon the theoretical analysis developped in (9) (10).

The paper is organized as follows. In section I the basic problem is stated. The section II is devoted to a (sketchy) theoretical derivation of the special cost-functions which have to be used instead of the classical mean-square error ; both real case (for one carrier systems) and complex case (for two carrier systems) are investigated ; the reader which is not interested in the theoretical analysis of the problem may drop this section. Blind equalization algorithms are presented in sections III (one-carrier transmission systems) and IV (two-carrier transmission systems). Experimental results are reported in section V, further ones are available in (11).

## I - STATEMENT OF THE PROBLEM.

The Fig. 1 and 2 show the general form of one-carrier amplitude modulation systems and two-carrier phase and amplitude modulation systems respectively, whereas the Fig. 2-a shows the complex baseband equivalent form of Fig.2.

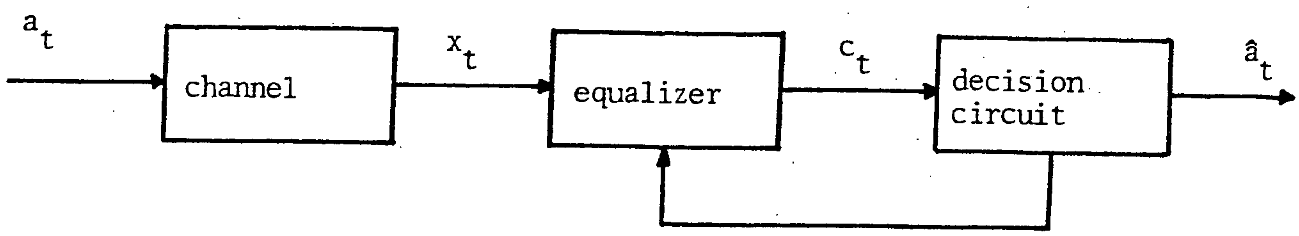



Fig. 1

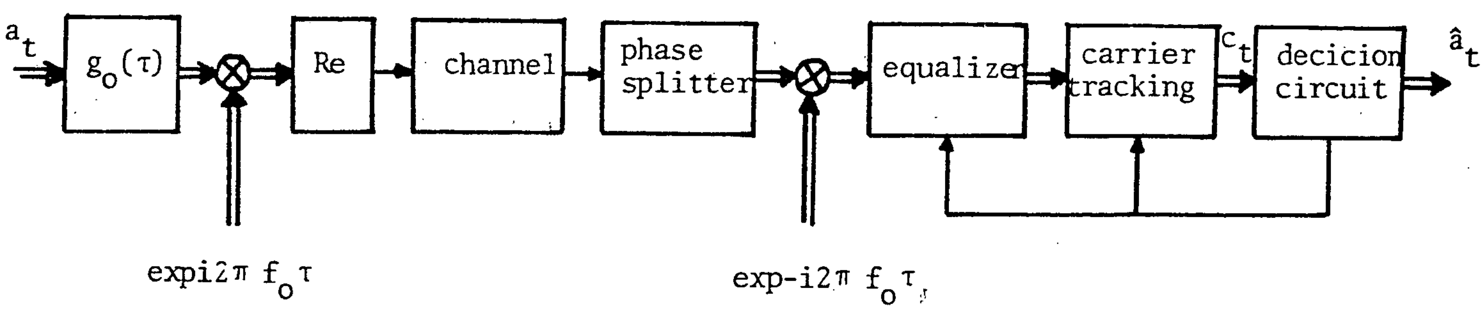



Fig. 2

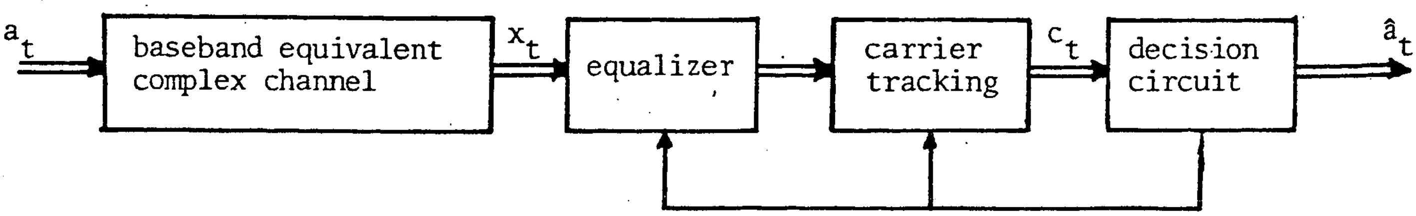



Fig. 2-a

For deriving standard equalization algorithms, a classical method is the following :

1) During the startup period, adjust the equalizer (and the phase estimator) in order to minimize the mean-square error

$$\mathbb{E} \quad |c_t - a_t|^2 ,$$

where $\mathbb{E}$ denotes the expectation over noise and data sequences, and $(a_t)$ is a known training sequence. Standard stochastic gradient (or "decision-directed") algorithms are used for this purpose.

2) For the transmission period, use the same algorithm as before, but replacing $a_t$ by its estimate $\hat{a}_t$. The so-obtained algorithms require a low error rate, and cannot be used during a starting period.

What we need for designing a blind equalizer is to recover the emited message $(a_t)$ from the received one $(x_t)$ only, without any prealable identification of the unknown channel. This requires that the input $c_t$ of the decision circuit be close to the true date $a_t$, which is unknown : this is the problem we shall solve in the sequel.

Let us point out that, for the purpose of the theoretical analysis, where the characteristics of the unknown channel (including its phase) are assumed to be time-invariant, the equalizer and phase estimator of Fig. 2-a are known to be redundant ( 5 ) ( 15 ) , so that the later will be dropped for the theoretical analysis in section II. Recall that splitting the two tasks of carrier recovery and removal of ISI is needed only because of the different time-scales of the time variations of the phase (fast) and the ISI (slow).

## II - GAIN AND PHASE IDENTIFICATION OF NONMINIMUM PHASE SYSTEMS.

### (II.1) The problem.

As it has been shown in section I, the theoretical problem we have to solve is the following : we observe the output $x_t$ of an unknown time-invariant linear system S admitting $a_t$ as input, where $a_t$ is an unobserved white (independent identically distributed) random sequence with known symmetric distribution $\nu$ of finite variance.

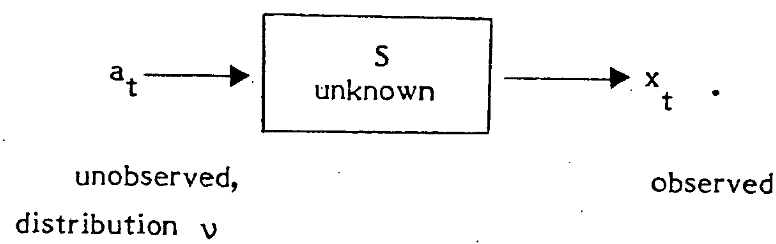



Fig. 3

PROBLEM : Reconstruct the message $a_t$, or, equivalently, identify the inverse $S^{-1}$ of the unknown channel.

The following remarks are fundamental :

1) There is no solution for this problem when the input $(a_t)$ is Gaussian (unless the channel S is known to be of minimal -or maximal- phase), since it is wellknown that only the attenuation of S can be recovered in this case, but not the group delay distorsion (i.e. phase of the linear system S). As a consequence, no solution can be found, which would be based upon second order statistics only.

2) Since S is assumed to be nonminimum phase, $S^{-1}$ is non causal. As a consequence the theoretical form of $S^{-1}$ will be, using the z-transform notation,

$$
\text{(II.1)} \quad \begin{cases} S^{-1}(z^{-1}) = \sum_{k \in \mathbb{Z}} \tilde{s}_k \, z^{-k}, & \text{where} \\ S^{-1} \circ S(z^{-1}) = z^{-N} & \text{for some } N, \end{cases}
$$

i.e. $S^{-1}$ is defined up to a time shift ; on the other hand, (II.1) implies that the message $(a_t)$ can be reconstructed in an offline way only. Obviously, in practical situations, $S^{-1}$ will be truncated, which will allow the reconstruction of $(a_t)$ in real time, but with a (constant and finite) delay.

In the sequel, unless the contrary is mentioned, $(a_t)$, $(x_t)$, and S will assumed to be real.

### (II.2) A unicity result.

The following result is proved in ( 9 ) :

THEOREM 1 : Assume $\nu$ is a nongaussian distribution. Let $\theta(z^{-1})$ be a linear system such that the random variable (for t arbitrary, but fixed)

$$
\text{(II.2)} \qquad c_t(\theta) \triangleq \theta(z^{-1}) \; S(z^{-1}) \; a_t
$$

be also of distribution $\nu$. Then,

$$
\text{(II.3)} \qquad \theta(z^{-1}) \; S(z^{-1}) = \pm \; z^{-N}
$$

for some N. Such a $\theta$ will be denoted by $\pm \; S^{-1}$.

In other words, if an equalizer (i.e. adjustable filter) is placed after the channel according to the following scheme,

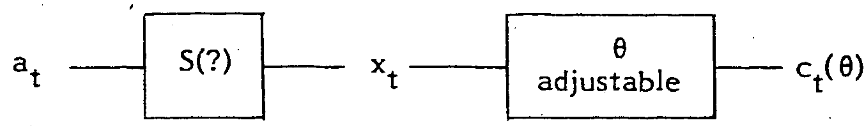



Fig. 4

it is sufficient to obtain that $c_t(\theta)$ admits $\nu$ as distribution, for ensuring that $\theta$ is, up to a sign, the desired inverse (the delay cannot be removed in view of (II.1)); the sign cannot be recovered since $\nu$ was assumed to be symmetric, so that nothing can distinguish the true message $(a_t)$ from the opposite one $(-a_t)$ from a statistical point of view. This unicity result is a very strong one, since nothing is requested about the joint distribution of the sequence $(c_t)$. Not surprisingly, the theorem is false when $\nu$ is gaussian!

As a consequence, it is not necessary to adjust $\theta$ by forcing the independence of the output $(c_t)$ of the equalizer, only the one-dimensional distribution of $c_t$ has to be looked for. The problem we shall solve now is the design of relevant cost functions for this purpose.

## (II.3) Gain-and-phase recovering cost funtions.

We were not able to design such cost functions for arbitrary $\nu$ ; let us define the family of distributions for which we will give the theoretical results :

DEFINITION : The distribution $\nu$ is said to be sub-gaussian in one of the following cases :

(II.4.i) $\nu$ is uniform over $[-d, +d]$,

(II.4.ii) $\nu(dx) = K \exp(-g(x))dx$,

where $K$ is a constant, and $g$ is an even function, derivable except possibly at the origin, such that $g(x)$ and $g'(x)/x$ are strictly increasing over $R_+$.

Examples : $\nu(dx) = K \exp(-|x|^{\alpha})$ for $\alpha > 2$, the limit case $\alpha = +\infty$ corresponds to the uniform distribution.

In view of the theorem 1, we look for cost functions of the form :

$$\text{(II.5)} \qquad J(\theta) = \mathbb{E}(\psi(c_t(\theta))),$$

where $\mathbb{E}$ denotes the expectation over all possible sequences of data ; the problem is to choose the function $\psi$. The following theorem is stated in a more general and precise way, and proved in [ 8 ] [ 9 ] :

THEOREM 2 : Assume $\nu$ is subgaussian. Then the Sato cost function, which is defined as follows,

$$\text{(II.6)} \qquad \begin{cases} J(\theta) = \mathbb{E}(\frac{1}{2} c_t^2(\theta) - \alpha \ |c_t(\theta)|) \\ \\ \alpha = \dfrac{\mathbb{E}\ a_t^2}{\mathbb{E}\ |a_t|} \end{cases},$$

admits as only local (and also global) minima the inverse channels $\pm S^{-1}$, as defined in theorem 1.

COMMENTS.

1/ This cost function was first used Sato [ 6 ] without theoretical justification. It is very interesting, since it needs the knowledge of only the first and second order moments of $\nu$, and leads to very simple algorithms.

2/ In [ 9 ], a family of suitable cost function was introduced, including the Sato function. Moreover, a way of solving the case of supergaussian distribution was given (the simplest example of supergaussian density is $\exp - |x|^{\beta}$ for $\beta < 2$).

3/ Let us give an explanation of the nice property of the subgaussian distributions we use in designing the cost functionals. Let us enforce the following energy constraint on the equalizer $\theta$

(II.7) $$\mathbb{E}\ c_t^2(\theta) = \mathbf{E}\ a_t^2 ,$$

which can be effectively realized in a simple way (see [9]). Assume $\nu(dx) = K \exp(-g(x))dx \overset{\Delta}{=} f(x)dx$ with g smooth. Then the following theorem is proved in [9] :

<u>THEOREM 3</u> : <u>Restricted to the manifold defined by the constraint</u> (II.7), <u>the cost functional</u>

(II.8) $$J(\theta) = \mathbb{E}(-\log f(c_t(\theta)))$$

<u>admits</u> $\pm\ S^{-1}$ <u>as only local minima.</u>

It can be shown that this functional is the most efficient one in the sense of the Cramer Rao bound and of the tracking capability of the corresponding identification algorithm [12].

Moreover, this function has an illuminating interpretation : denoting by $\nu_c$ the distribution of $c_t(\theta)$, and by $d\nu_c/dx$ its density with respect to the Lebesgue measure, we get

(II.9)
$$\begin{aligned} J(\theta) &= -\int \log \frac{d\nu}{d\nu_c} \cdot d\nu_c - \int \log \frac{d\nu_c}{dx}\, d\nu_c \\ &= H(\nu_c/\nu) + H(\nu_c) , \end{aligned}$$

where $H(\nu_c/\nu)$ is the Kullback Information of $\nu_c$ with respect to $\nu$, and $H(\nu_c)$ the selfentropy of $\nu_c$. Since it is known that the Kullback Information is a good measure of distance between two distributions for ensuring the desired behaviour for $J(\theta)$ in (II.9), it is sufficient to know that <u>the self entropy of $\nu_c$ is an increasing function of the distance of the global channel</u> $\theta(z^{-1}) \cdot S(z^{-1})$ <u>to</u> $\pm$ Identity.

But this is exactly the kind of nice property the subgaussian distribution enjoy.

4/ <u>Finding the good sign in (II.3).</u>

The figure 5 shows the steepest descent lines of $J(\theta)$ for the case where the global channel $T = \theta.S$ is subject to have only two non zero coefficients.

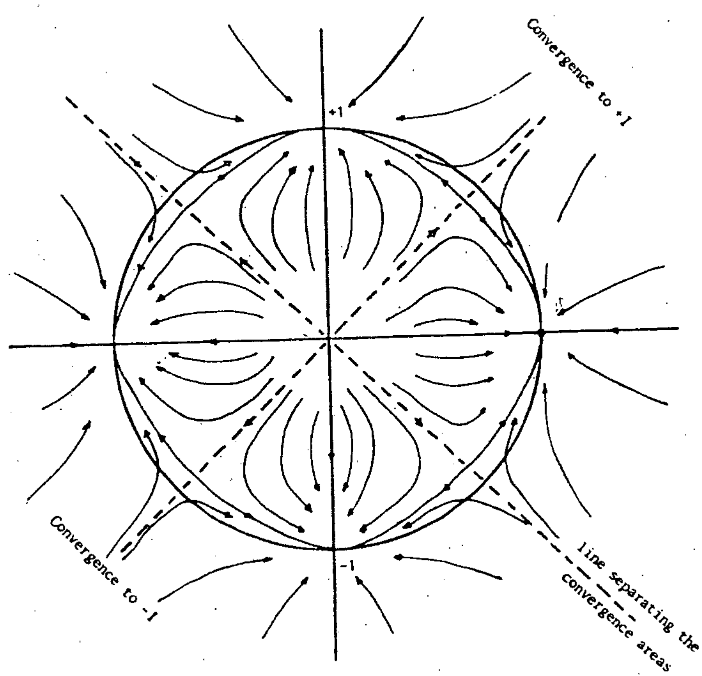



Fig. 5

It appears clearly that, using a steepest descent method for minimizing J, it is sufficient to initialize the algorithm in the correct half space for $\theta$ ; and it turns out that the very coarse information on S which is needed for a good initialization is always available (in practice $\theta = +$ Identity is always a good initialization).

5/ use on discrete distributions.

The theorem we have given is not entirely satisfactory in our case, since the distributions $\nu$ encountered in data transmission are not subgaussian in our sense. Such typical distributions are uniform distributions over finite sets of the form $\{\pm 1, \pm 3, \ldots, \pm (2K+1)\}$ ; such distributions are "close" to the uniform distribution, at least for K not too small ; furthermore, they enjoy the nice property explained after theorem 3, namely the self entropy of $c_t(\theta)$ subject to $\mathbb{E}\, c_t^2 = \mathbb{E}\, a_t^2$ increases with the distance of $\theta.S$ from the identity. Although we have no proof for these situations, the following simple example will illustrate what happens.

EXAMPLE : Let $a_t$ be +1 , -1 with probability 1/2, and consider $c_t(\theta) = \theta(z^{-1})\, S(z^{-1}) a_t$, subject to the constraint $\mathbb{E}\, c_t^2 = \mathbb{E}\, a_t^2 = 1$, in the two following particular cases :

Case 1 : $\theta$ is such that the global channel $T = \theta.S$ has only two non zero coefficients cos x and sin x for $0 \le x \le \pi/4$ (which is sufficient for reasons of symmetry), so that

$$c_t(x) = \cos x . a_t + \sin x . a_{t-1} .$$

Taking for J the Sato cost function, we get, since $\alpha = 1$,

$$J(\theta) \stackrel{\Delta}{=} J(x) = \mathbb{E}(c_t^2(x) - 2|c_t(x)|)$$

$$= 1 - 2\, \mathbb{E}\, |c_t(x)| \qquad \text{(II.10)}$$

$$= 1 - 2 \cos x ,$$

which is increasing with x for $0 \le x \le \pi/4$ ; this is the result we need, since the identity corresponds here to $x = 0$.

Case 2 : $\theta$ is such taht the global channel is far from the identity. In this case we can use the central limit theorem, and consider that $c_t$ is approximately Gaussian (0,1), so that

(II.11)
$$J(\theta) = 1 - 2(2\pi)^{-1/2} \int_{-\infty}^{+\infty} |x| \, e^{-\frac{x^2}{2}} \, dx = 1 - 4. \, (2\pi)^{-1/2}$$
$$> - 1 = J(S^{-1}),$$

so that J is well conditionned far from the solution.

Although non theoretically proved, this good behaviour is confirmed by the experimentations.

Rather than a comprehensive theoretical treatment of our problem, the results we have presented in this section have to be considered as a guide for designing gain and phase recovering cost functions.

## (II.4) Extension to the complex case.

Here we consider the case where the signals $a_t$, $x_t$, $c_t$ and the channel S and equalizer $\theta$ are complex.

THEOREM 4 : Assume we are in the case of two independent carrier, namely the message $(a_t)$ is an independent identically distributed (i,i.d.) sequence such that Re $a_t$ and Im $a_t$ be independent, with same subgaussian distribution $\nu$ . Then the complex Sato cost function

(II.12)
$$\begin{cases} J(\theta) = \mathbb{E}\,(\psi(\mathrm{Re}\ c_t(\theta)) + \psi\ (\mathrm{Im}\ c_t(\theta))) \\ \psi(x) = \frac{1}{2}\, x^2 - \alpha\,|\,x\,| \qquad \text{(Sato function)} \\ \alpha = (\int x^2\ \nu(dx))\ (\int |x|\ \nu(dx))^{-1} \end{cases},$$

admits $\pm\, S^{-1}$ as only local minima.

PROOF : Define for $t \in Z$

(II.13) $$b_{2t} = \text{Re } a_t \quad , \quad b_{2t+1} = \text{Im } a_t \; .$$

Then $(b_t)$ is a real valued i.i.d. sequence with subgaussian distribution $\nu$, and the random variable $\text{Re } c_o(\theta)$ is of the form

(II.14) $$\text{Re } c_o(\theta) = \sum_{k \in \mathbb{Z}} t_k \, b_{-k}$$

for some real coefficients $(t_k)$ depending upon $\theta$. By theorem 2, the local minima of the cost function

(II.15) $$\mathbb{E}(\psi(\text{Re } c_o(\theta)))$$

correspond to those $T(z^{-1}) = \sum_t t_k \, z^{-k}$ equal to $\pm z^N$ for some delay N. Hence, according N be odd or even,

(II.16) $$\begin{cases} \mathbb{E}\,(\psi(\text{Re } c_o(\theta))) \text{ is minimum for} \\ \\ \text{Re } c_t(\theta) = \begin{cases} \pm \text{ Re } a_t \\ \text{or} \\ \pm \text{ Im } a_t \end{cases} \quad , \text{ up to a delay,} \end{cases}$$

which characterizes the desired inverse $S^{-1}$ in the complex case. The same as in (II.16) holds with $\text{Im } c_o(\theta)$ instead of $\text{Re } c_o(\theta)$, thus leading to another convenient cost function. Add both of them for a better efficiency, this result in the Sato cost function of the formula (II.12), which proves the theorem.

This result will be extremely useful for designing blind equalizers in the two carrier case. The same remarks hold as in the real case. Unfortunately, we have no similar result for the case of two dependent carrier (as it is the case when the 16 -point V29 CCITT- constellation is used, see figure 18 ), which will result in great difficulties in designing blind equalizers in this case (Recall the weaker results of [7] do not suffer from this restriction).

## III - DESIGN OF BLIND EQUALIZERS FOR ONE-CARRIER TRANSMISSION SYSTEMS

We shall use the Sato cost function (and modifications of them) for the case of $(a_t)$ being uniformly distributed over the set $\{\pm 1, \pm 3, \ldots \pm (2K+1)\}$, which corresponds to Amplitude modulation schemes.

### III.1 - General form of the algorithm

We shall use stochastic gradient methods, and self orthogonalizing algorithms for minimizing the cost functions we have defined ; a theoretical analysis of the convergence of such adaptive schemes is given in (10).

The general form of the cost function we shall use is the following

(III.1) $$J(\theta) = \mathbb{E}(\psi(c_t(\theta))) \; ;$$

for example, the cost function of theorem 2 corresponds to the choice of the Sato function (II.12) for $\psi$.

Defining :

(III.2) $$\begin{cases} \frac{\partial}{\partial\theta} c_t(\theta) = X_t(\theta) \\ \varepsilon_t(\theta) = -\psi'(c_t(\theta)) \end{cases}$$

($\psi'$ denoting the derivation of $\psi$), where $\varepsilon_t(\theta)$ will be referred to as the pseudo-error signal, we get :

(III.3) $$\text{grad } J(\theta) = -\mathbb{E}\,(X_t(\theta)\ \varepsilon_t(\theta)).$$

These considerations lead to the following general form for the corresponding stochastic gradient algorithm ((9), (10)) :

(III.4) $$\theta_{t+1} = \theta_t + \gamma X_t(\theta_t)\ \varepsilon_t(\theta_t),$$

where $\gamma$ is some small gain. Self-orthogonalizing procedures can also be designed according to the particular cases.

The following points have now to be investigated :

a) How to synthetize the equalizer $\theta$ ?

The theoretical form $\theta(z^{-1}) = \sum_{k \in \mathbb{Z}} \theta_k z^{-k}$ cannot be used in the practical algorithms, since only finitely many parameters can always be used. The effect of such a truncation on the recovering of $S^{-1}$ is investigated in [ 8 ], [ 9 ]. The two classical transversal and recursive structures will be used for synthetizing $\theta$.

b) How to choose the pseudo-error signal $\varepsilon_t(\theta)$ ?

The Sato cost function suggests a choice for $\varepsilon_t$, but other will be investigated, in order to improve the properties of the blind equalizer.

Note that both points a) and b) are in fact disconnected : any relevant pseudo-error signal can be used with any of the two considered structures. Let us now investigated in more details these points.

III.2 - Some pseudo-error signals

Let us explain the Sato pseudo-error signal, which corresponds to $\psi$ as in (II.12) :

(III.5) $$\varepsilon_t^S(\theta) = \alpha.\text{sgn } c_t(\theta) \ , \ \alpha = \frac{\mathbb{E}\, a_t^2}{\mathbb{E}|a_t|} .$$

Here, $\alpha.\text{sgn } c_t$ plays the rôle of a coarse estimate of the true signal $a_t$, hence the name of "pseudo-error". Unfortunately, the behavior of this pseudo-error signal around the solution ($\theta \simeq S^{-1}$) is very noisy (unless $a_t$ takes only the values $\pm$ 1), since, although being zero-mean, the signal $\varepsilon_t^S(\theta)$ is still non-zero for $\theta = S^{-1}$.

On the other hand, the customary pseudo-error signal :

(III.6) $$e_t(\theta) = \hat{a}_t(\theta) - c_t(\theta)$$

(where $\hat{a}_t$ is the estimate of $a_t$ based upon $c_t$), which is used in the standard selfadaptive equalizers, is not robust, but enjoys the desirable property of being zero for $\theta = S^{-1}$.

Hence the idea of combining both signals for obtained the following G-pseudo-error signal :

(III.7) $$\varepsilon_t^G(\theta) = k_1 \; e_t(\theta) + k_2 |e_t(\theta)| . \varepsilon_t^S(\theta) \; ,$$

where $k_1$ and $k_2$ are constants. The behavior of this G-pseudo-error signal is the following. For $\theta$ being far from $S^{-1}$, $|e_t(\theta)|$ is large, and the second term ensures the robustness of the blind equalizer ; on the other hand, for $\theta$ close to $S^{-1}$, the second term has the same order of magnitude as the first one and $\varepsilon_t^G(\theta) = 0$ for $\theta = S^{-1}$, which ensures the removal of the noise due to $\varepsilon_t^S$.

Hence the G-pseudo-error signal provides us with a smooth automatic switching from a blind startup period to the conventional equalization mode ; conversely, in case of an abrupt change in the characteristics of the unknown channel S, this signal goes back automatically to the blind startup mode. No further testing is needed for such switches. This is a highly interesting property of this kind of pseudo-error signal.

III.3 - Structure of the equalizer

Transversal Equalizer

Here $\theta$ is synthetized in the standard transversal form. The corresponding algorithm is

(III.8) $$\begin{cases} \theta_{t+1} = \theta_t + \gamma \, X_t . \varepsilon_t(\theta_t) \\ X_t^T = (x_{t+N}, \ldots, x_t, \ldots, x_{t-N}) \\ c_t(\theta) = \sum_{-N}^{+N} \theta_k . x_{t-k} = X_t^T . \theta \; , \end{cases}$$

where $\varepsilon_t(\theta)$ is one of the pseudo-error signals which was given in the preceding section.

A self-orthogonalizing procedure is suggested in ( 9 , section F ), and was experimented in ( 8 ). However multistep procedures are nearly as efficient, and require much less computational effort (11 ), even compared with fast algorithms in ladder form.

Recursive Equalizer

Here, $\theta$ is synthetized in the recursive form,

$$
\text{(III.9)} \quad \left\{
\begin{aligned}
c_t(\theta) &= \sum_{+M}^{+N} \theta_k^1 \, x_{t-k} + \sum_{1}^{P} \theta_k^r \, c_{t-k}(\theta) \triangleq Z_t^T(\theta) \, . \, \theta \ , \\
\theta^T &= (\theta_{-M}^1 \ ,\dots,\ \theta_{+N}^1 \ ;\ \theta_1^r \ ,\dots,\ \theta_P^r) \\
Z_t^T(\theta) &= (x_{t+M} \ ,\dots,\ x_{t-N} \ ;\ c_{t-1}(\theta) \ ,\dots,\ c_{t-P}(\theta)) \ ;
\end{aligned}
\right.
$$

here no decision is done in the loop. It is proved in ( 14 ) that

$$
\text{(III.10)} \quad
\begin{aligned}
&\frac{\partial}{\partial \theta} c_t(\theta) = X_t(\theta) \ , \text{ where} \\
&X_t(\theta) = \frac{1}{1 - \sum_{1}^{P} \theta_k^r \, z^{-k}} \, Z_t(\theta) \ ,
\end{aligned}
$$

so that the "state vector" $X_t(\theta)$ is the output of the recursive part of the filter $\theta$, with $Z_t(\theta)$ as input. The corresponding algorithm is given in (III.4) ; he is known to converge to the desired value of $\theta$, provided that it does not become unstable, which can be always guaranteed using simple modifications of this algorithm (13) ; when the classical least squares cost function is used instead of the present one, this procedure is known as the stochastic approximation version of the Recursive Maximum Likelihood procedure (14 ).

However this procedure, which involves the filtering of the "state vector", is never used in the classical recursive equalizers. The counterpart of the classical recursive structure (without decision in the loop ! ) is here the following :

$$
\text{(III.11)} \quad \left\{ \begin{aligned}
&\theta_{t+1} = \theta_t + \gamma Z_t \, \varepsilon_t \,, \quad \varepsilon_t = -\psi'(c_t) \,, \\
&Z_t^T = (x_{t+M} ,\dots, x_{t-N} \,;\, c_{t-1} ,\dots, c_{t-P}) \\
&c_t = \sum_{-M}^{+N} \theta_k^1(t) \, x_{t-k} + \sum_{1}^{P} \theta_k^r(t) \, c_{t-k} \,,
\end{aligned} \right.
$$

i.e. the state vector $Z_t$ is used without any filtering in the time-update recursion of the blind equalizer. It is known that such a simplification is not guaranteed to converge in any situation (14) ; it is proved in (13 ) that a sufficient condition for the desired value $\theta_*$ of the equalizer to be a stable equilibrium of the algorithm is that the transfer function

$1 - \sum_{1}^{P} \theta_{*k}^r \, z^{-k}$ be <u>positive real</u>, i.e.

$$
\text{(III.12)} \qquad \mathrm{Re}(1 - \sum_{1}^{P} \theta_{*k}^r \, e^{ikw}) > 0 \text{ for } - \pi < \omega < \pi \quad ,
$$

a condition which is satisfied by most of the channels. In any case the recursive algorithm (III.11) behaves differently as the transversal one which corresponds to a gradient method ; although it can be used when a recursive form is desired for the equalizer during the standard reception period, this structure is less efficient than the transversal one during the blind equalization period, as shown by the experiments.

## IV - DESIGN OF BLIND EQUALIZERS FOR TWO CARRIER TRANSMISSION SYSTEMS.

Here we are interested in joint amplitude and phase modulation schemes ; the results on the case where $a_t = \pm 1$ (section II.3) suggest that the classical error signals are suitable for blind equalization when pure phase modulation is used, which was also pointed out in ( 7 ), and is verified by the experiments.

Starting from the cost functions we have developped for the complex case, we shall consider two situations.

a) The equalization is performed independently from the carrier recovery ; provided that we work in the baseband equivalent form, the corresponding algorithms will be the exact counterpart of the former ones to the complex case.

b) Joint equalization and carrier recovery is performed, in the spirit of ( 15 ) ( 5 ). This is useful in case the time variations of the phase of the channel (jitter, offset,...) are fast compared to other time variations of the channel (corresponding to the intersymbol interference). The theoretical analysis we have done will provide us with blind equalizers which are much more efficient than those presented in ( 7 ).

The theoterical derivation of the algorithms will be done in the two independent carrier case only, since no similar theoretical approach is available in the general case ; other examples, like the V.29 constellation of the CCITT will be considered via ad hoc techniques, without theoretical investigation.

### IV.1 - General form of the blind equalizer in baseband equivalent form, when no joint carrier recovery is done.

The corresponding scheme is

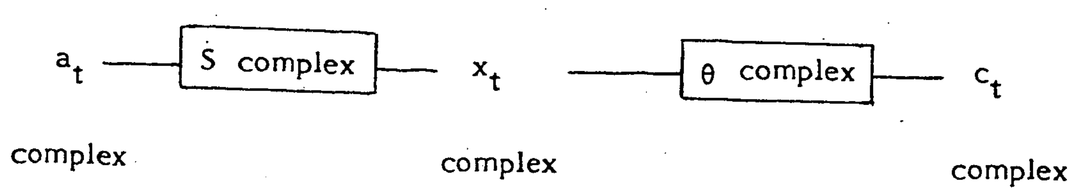



FIG. 6

Recall the cost functions we shall use are of the form

$$\text{(IV.1)} \qquad J(\theta) \doteq \mathbb{E}\,(\psi(\mathrm{Re}\ c_t(\theta)) + \psi(\mathrm{Im}\ c_t(\theta)))\ ,$$

where $\psi$ is one of the functions introduced in section II. Setting

$$\text{(IV.2)} \qquad \mathrm{grad}\ J \overset{\Delta}{=} \frac{\partial}{\partial(\mathrm{Re}\ \theta)} J + i \frac{\partial}{\partial(\mathrm{Im}\ \theta)} J\ ,$$

a straightforward calculation leads to

$$\mathrm{grad}\ J(\theta) = -\ E\,(X_t^*(\theta)\ \varepsilon_t\ (\theta))\ ,^{(1)} \qquad \text{where}$$

$$\text{(IV.3)} \qquad \begin{aligned} X_t(\theta) &\overset{\Delta}{=} \frac{\partial}{\partial(\mathrm{Re}\ \theta)} (\mathrm{Re}\ c_t(\theta)) + i \frac{\partial}{\partial(\mathrm{Re}\ \theta)} (\mathrm{Im}\ c_t(\theta)) \\ &= \frac{\partial}{\partial(\mathrm{Im}\ \theta)} (\mathrm{Im}\ c_t(\theta)) - i \frac{\partial}{\partial(\mathrm{Im}\ \theta)} (\mathrm{Re}\ c_t(\theta))\ , \end{aligned}$$

$$\varepsilon_t(\theta) \overset{\Delta}{=} -\ (\psi'(\mathrm{Re}\ c_t(\theta)) + i\ \psi'(\mathrm{Im}\ c_t(\theta)))\ ,$$

where the two different expressions for $X_t(\theta)$ are in fact equivalent because $c_t(\theta)$ is an holomorphic function which satisfies consequently the Cauchy-Rieman equation.

This general formula for the gradient would enable us again to derive algorithms with both transversal and recursive structures for the equalizer. For the sake of simplicity, we shall restrict ourselves to the transversal case (see [11] for results on recursive structures). According to the formulas (IV.3), the corresponding stochastic gradient algorithm is the following

$$\text{(IV.4)} \qquad \begin{cases} \theta_{t+1} = \theta_t + \gamma\ X_t^*\ \varepsilon_t(\theta_t) \\ X_t^T \overset{\Delta}{=} (x_{t+N}, \ldots, x_{t-N}) \\ c_t(\theta) = \sum\limits_{-N}^{+N} \theta_k\ x_{t-k} = X_t^T \cdot \theta \end{cases}$$

where the pseudo-error signal has to be chosen.

(1) the superscript * denotes the complex conjugate.

(IV.2) - General form of the equalizer when joint blind equalization and carrier recovery is performed.

Without loss of generality, we can assume that demodulation by a suitable local carrier (without phase tracking !) is carried out before equalization and phase recovery, so that we can investigate the problem in the baseband equivalent form ([2] [3] [5]). We shall essentially follow the argument of [5] [15] for deriving equalizers with joint carrier recovery.

The equalizer is splitted in two parts, yielding

(IV.5) $$c_t(\theta,\phi) = c_t(\theta)\, e^{-i\phi} \quad ,$$

where the purpose of $\theta$ is to remove the intersymbol interference (which is assumed to be "slowly" varying), whereas the phase parameter $\phi$ will be devoted to the tracking of "fast" tap rotation effects (phase jitter, frequency offset,...) ; as it has been pointed out by D.D. Falconer [5], these two parts are redundant in the theoretical case of a time invariant channel with time invariant phase.

With the cost function

(IV.6) $$J(\theta,\phi) \triangleq \mathbb{E}\,(\psi(\mathrm{Re}\ c_t(\theta,\phi)) + \psi(\mathrm{Im}\ c_t(\theta,\phi))) \quad ,$$

and setting

(IV.7) $$\mathrm{grad}_\theta\, J \triangleq \frac{\partial}{\partial(\mathrm{Re}\ \theta)} J + i \frac{\partial}{\partial(\mathrm{Im}\ \theta)} J \quad ,$$

we get, as previously

$$\mathrm{grad}_\theta\, J(\theta,\phi) = -\,\mathbb{E}\,(X_t^*(\theta,\phi)\;\varepsilon_t(\theta,\phi)) \quad ,$$

$$\text{(IV.8)}\qquad X_t(\theta,\phi) \triangleq \frac{\partial}{\partial(\mathrm{Re}\;\theta)}\,(\mathrm{Re}\;c_t(\theta,\phi)) + i\,\frac{\partial}{\partial(\mathrm{Im}\;\theta)}\,(\mathrm{Im}\;c_t(\theta,\phi))$$

$$\varepsilon_t(\theta,\phi) = -\,(\psi'(\mathrm{Re}\;c_t(\theta,\phi)) + i\;\psi'(\mathrm{Im}\;c_t(\theta,\phi))) \quad ,$$

whereas

$$\text{(IV.9)}\qquad \frac{\partial}{\partial\phi} J(\theta,\phi) = -\;\mathrm{Im}\;(c_t(\theta,\phi)\;\varepsilon_t^*(\theta,\phi)) \;.$$

Restricting ourselves to the transversal structure for the equalizer, we get the following algorithm

$$\text{(IV.10)}\qquad \begin{cases} \theta_{t+1} = \theta_t + \gamma\, X_t^*\, e^{+i\phi_t}\, \varepsilon_t(\theta_t,\phi_t) \\ \phi_{t+1} = \phi_t + \mu\, \mathrm{Im}(c_t(\theta_t,\phi_t).\;\varepsilon_t^*(\theta_t,\phi_t)) \\ X_t^T = (x_{t+N}, \ldots, x_{t-N}) \\ c_t(\theta,\phi) = X_t^T.\theta\; e^{-i\phi} \;, \end{cases}$$

whereas the pseudo-error signal $\varepsilon_t$ has to be chosen.

## IV.3 - Choice of the pseudo-error signal.

Here we shall choose the signal $\varepsilon_t$ in order to ensure the robustness of the equalizer during a blind startup period. We restrict ourselves to the case of two independent carriers, where the real and imaginary parts of the message take the values $\{\pm 1, \pm 3, \ldots, \pm 2K+1\}$ with equal probability. Some pseudo-error signals for the 16-point V.29 constellation of the CCITT are given in the fig. 18.

Let us begin with the complex extension of the Sato cost function, which corresponds to the formulas (IV.3) or (IV.8) with $\psi$ given in (II.12). This gives, with dropping the obvious dependence on $\theta$ and $\phi$,

$$\varepsilon_t^s = c_t - \alpha\hat{c}_t \text{ , where}$$

(IV.11) $$c_t \triangleq a\ (\operatorname{sgn}(\operatorname{Re} c_t) + i \operatorname{sgn}(\operatorname{Im} c_t))$$

$$\alpha = (\int x^2\, \nu(dx))\ (\int x\, \nu(dx))^{-1} \quad ,$$

where $\nu$ is the (common) distribution of Re $a_t$ and Im $a_t$. Note that, in the case of the 4-phase modulation scheme, $\hat{c}_t$ is nothing but the message reconstructed at the receiver, namely $\hat{a}_t$, so that $\varepsilon_t^s$ is identical to the classical error signal, which turns out to be robust in this case. In the general case, it appears that $\hat{c}_t$ is nothing but the best estimate among the possible values of an equivalent 4-phase constellation. The figure 7 shows the resulting constellaion for the case of a QAM-16 modulation scheme.

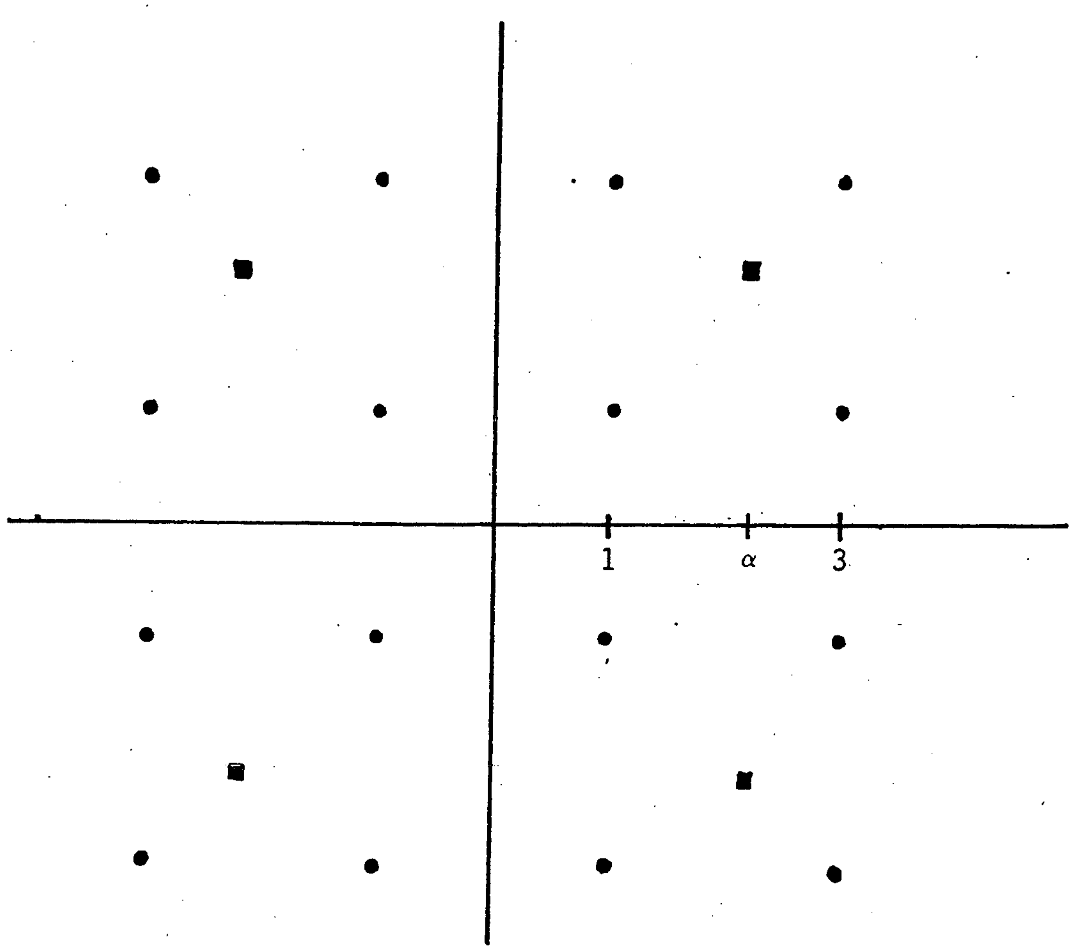



Fig. 7

As in the one-carrier case, it is possible to remove the noise due to the use of $\varepsilon_t^S$ even in the case of perfect equalization, since $\varepsilon_t^S$ can never be zero. Denoting by $\hat{a}_t$ the signal which is reconstructed from $c_t$ using the standard decision rules, let us introduce the modified pseudo-error signal

$$\text{(IV.12)} \quad \begin{cases} \varepsilon_t^G = k_1 \; e_t + k_2 \; |e_t| \; \varepsilon_t^S \text{ , where} \\ e_t \triangleq \hat{a}_t - c_t \text{ .} \end{cases}$$

As discussed before, this pseudo-error signal is again robust, allows the removal of the noise due to $\varepsilon_t^S$ at the equilibrium, and provides us with a (smooth) automatic switching between the blind startup period and the normal receiving period (and vice-versa in case of an abrupt change in the characteristics of the channel).

REMARK : a simplification of the loop error signal.

The two pseudo-error signals $\varepsilon_t^S$ and $\varepsilon_t^G$ have the property that

$$\text{(IV.13)} \quad \begin{cases} \operatorname{Im} \; (c_t(\varepsilon_t^S)^*) = \operatorname{Im} \; (c_t \; \hat{c}_t^* ) \\ \operatorname{Im} \; (c_t(\varepsilon_t^G)^*) = \operatorname{Im} \; (c_t(k_1 \; \hat{a}_t^* \; + k_2 \, |e_t| \, \hat{c}_t^* )) \text{ ,} \end{cases}$$

where $\hat{c}_t$ is defined in (IV.11) ; hence the loop error signals are of the same form as in the case of a Decision Feedback loop for a 4-Phase Shift Keying modulation. This strongly suggests a 4-th power loop could be used instead of the loop given in (IV.10).

Summary of the algorithms for blind joint equalization and carrier recovery.

Using the pseudo-error signal (IV.11) of Sato type.

For the case of a transversal equalizer, the algorithm is the following (see (II.12) for the definition of $\alpha$) :

$$
\text{(IV.14)} \quad \left\{ \begin{aligned}
\theta_{t+1} &= \theta_t + \gamma\, X_t^* \, e^{i\phi_t} \, \varepsilon_t^s \\
\phi_{t+1} &= \phi_t + \mu \, \mathrm{Im}(c_t \, \hat{c}_t^*) \\
c_t &= X_t^T \, \theta_t \, e^{-i\phi_t} \, , \quad \varepsilon_t^s = \hat{c}_t - c_t
\end{aligned} \right.
$$

$$
\hat{c}_t = \alpha \, (\mathrm{sgn}\, (\mathrm{Re}\; c_t) + i \; \mathrm{sgn}\, (\mathrm{Im}\; c_t)) \, ,
$$

which corresponds to the following block-diagram :

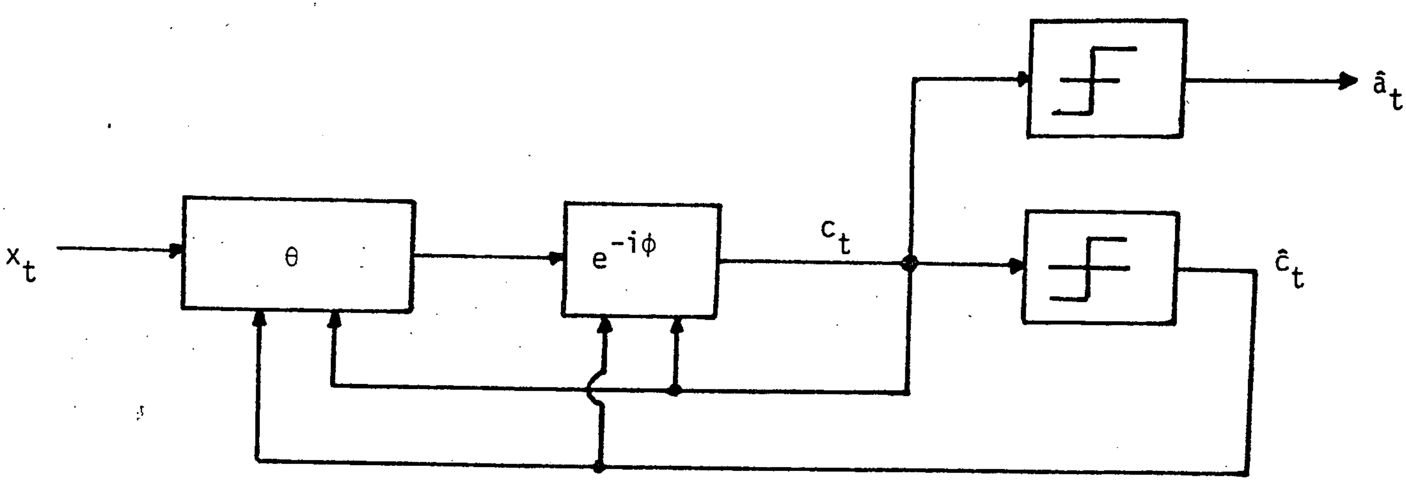



Fig. 8

Using the pseudo-error signal (IV.12) with smooth switching between the blind startup and normal periods.

The algorithm is the following

$$
\text{(IV.15)} \quad \begin{cases} \theta_{t+1} = \theta_t + \gamma\, X_t^* e^{i\phi_t}\, \varepsilon_t^G \\ \phi_{t+1} = \phi_t + \mu\, \mathrm{Im}(c_t\, \varepsilon_t^{G*}) \\ \varepsilon_t^G = k_1\, e_t + k_2\, |e_t|\, \varepsilon_t^s , \end{cases}
$$

where $e_t$ is the standard error signal defined in (IV.12), whereas $\varepsilon_t^s$ is defined in (IV.14). The corresponding block-diagram is the following

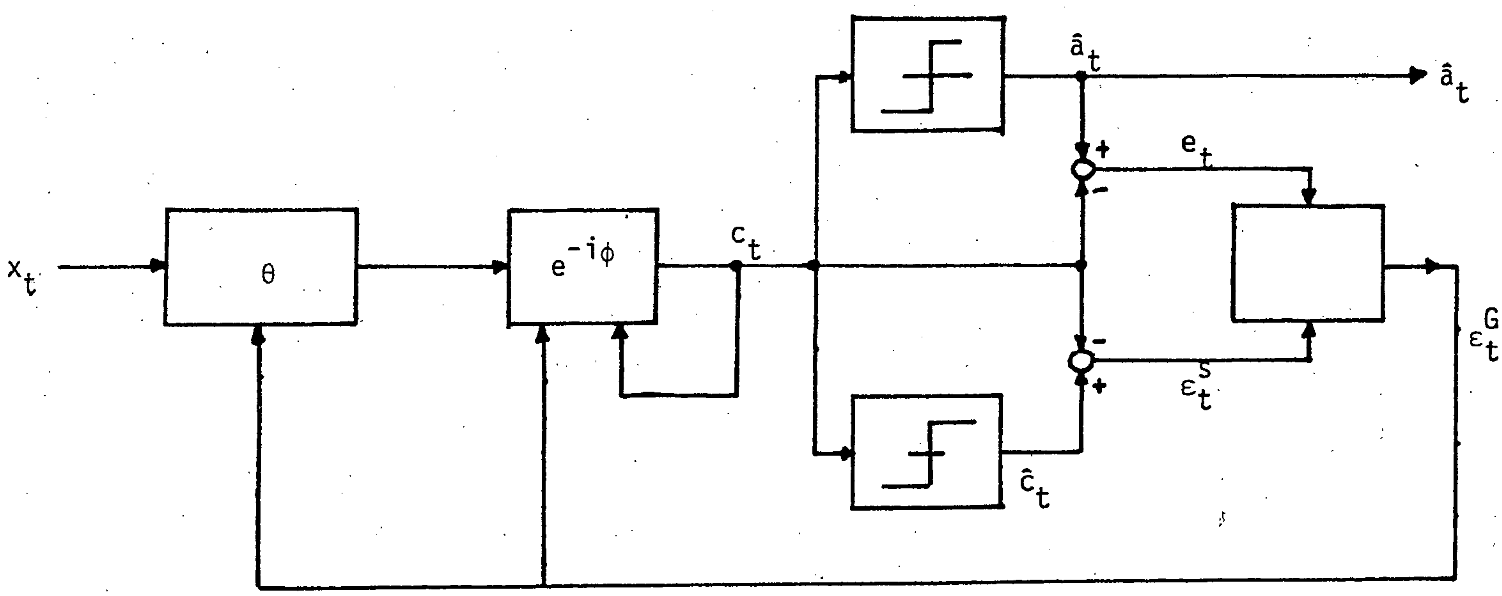



Fig. 9

As a conclusion of this section, let us emphasize that the theoretical analysis we have done provides us with tools for deriving various forms for a blind equalizer, among them only the most illustrative were given here. Finally let us mention the case of some schemes where the two carrier are dependent : in the fig. 18 , a blind equalizer is given for the

V.29 constellation of the CCITT ; this equalizer was derived using ad hoc arguments, and is less efficient than those designed for QAM constellations, but still more efficient than those proposed in ( 7 ).

# V - COMPUTER SIMULATIONS

## (V.1) One carrier

The system used is the amplitude modulation. The transmitted data are equally distributed random variables on the 8 levels set $\{\pm 1 \pm 3 \pm 5 \pm 7\}$. The impulse response of the channel is given by the figure 10 : the chosen sample is $1/3200_s$ giving an output of 9600 bits/s. The additive noise on the output is given by simulation of a gaussian r.v. $\mathcal{N}(0,\sigma^2=0.02)$.

The number of the tap weights for the equalizer is n=21. The algorithms are :

AN1 given by III.4 with the pseudo error III.5

ASS1 given by III.4 with the pseudo error IV.7.

For AN1 $\gamma = 10^{-4}$ and for ASS1 $\gamma = 2.10^{-4}$

The figure 11 gives the evolution of the error rate and the corresponding mean squared error for AN1 and the Fig. 12 the same results for ASS1.

The duration of the blind training phase is here about 0.5s. With a slightly modified algorithm (relaxation) this duration becomes 300 ms.

We can mention some other experiments with good results : different numbers of tap weights for the equalizer, insertion of a whitening filter before the equalizer, drastic changes of the impulse response (cf [11]).

## (V.2) Two carriers

### a) With no carrier tracking

In this case the algorithms are :

AN2 given by IV.14

ASS2 given by IV.15.

The impulse responses are given by the figure 13 corresponding to the amplitude and delay distorsions given by the figure 14 The output is 9600 bits/s. The additive white noise on the output (as for the following examples) is given by the simulation of two independent gaussian r.v. $N(0,\sigma^2=0.02)$. The equalizer is with a double sampling : n = 21 tap weights for the "main sample" and 11 tap weights for the shifted sample.

For the MAQ transmission the results corresponding to AN2 and ASS2 are respectively given by the Figure 15 and 16.

The results of ASS2 for the amplitude and phase modulation corresponding to the 16 points V29 constellation are given by the figure 17

In this last case the decision areas for the received data are given by the figure 18.

b) Jointly self recovering equalization and carrier tracking

- Case 1 frequency offset

We consider the same channel as in the previous case (figures 13 and 14 with the MAQ transmission and we add a 7Hz frequency offset. The corresponding results for the algorithm ASS2 are shown on the figure 19.

- Case 2 frequency offset and change of the channel.

Consider a line $S = \{S_i\}_{-N \leq i \leq N}$ and

$$t(S) = T = \{t_i\} \text{ with } \begin{cases} t_i = -0.5 . s_i & \text{for } i \neq 0 \\ t_0 = s_0 \end{cases}$$

We start with T = t(S), S given by the figure 13 and a 2Hz frequency offset ; after 2000 data T switches to S with keeping the same offset. The results are on the figure 20.

Remark

The computer simulations for different other cases are given in (11)

- simplified pseudo error signal $\varepsilon_t \rightarrow$ sign $(\varepsilon_t)$
- relaxation over the tap weights in the stochastic approximation algorithm.
- using a better integration method (multistep stochastic gradient)
- drastic channel distorsion.

## VI. CONCLUSIONS

The algorithms which we have proposed here seem to have essential good properties :

- same implementation and same computer power than the classical stochastic gradient algorithm for minimizing the mean-squared error.

- same form for all the different transmission systems : one or two carriers, MAQ, V29 ...

- convergence of the equalizer of the order of 1s

- automatic smooth switching to an algorithm quite identical to the classical one with possible automatic restarting.

- extrême robustness with respect to the number of digits used in computing shown by the results given in (11) when $\varepsilon_t$ or $X_t$ are replaced by sign $\varepsilon_t$ and sign $X_t$

- possible modifications in order to increase the performance if necessary (case 12000 bits/s) : relaxation in the adjustment of the tap weights ($n \rightarrow 2n$ multiplications) and more precise integration method.

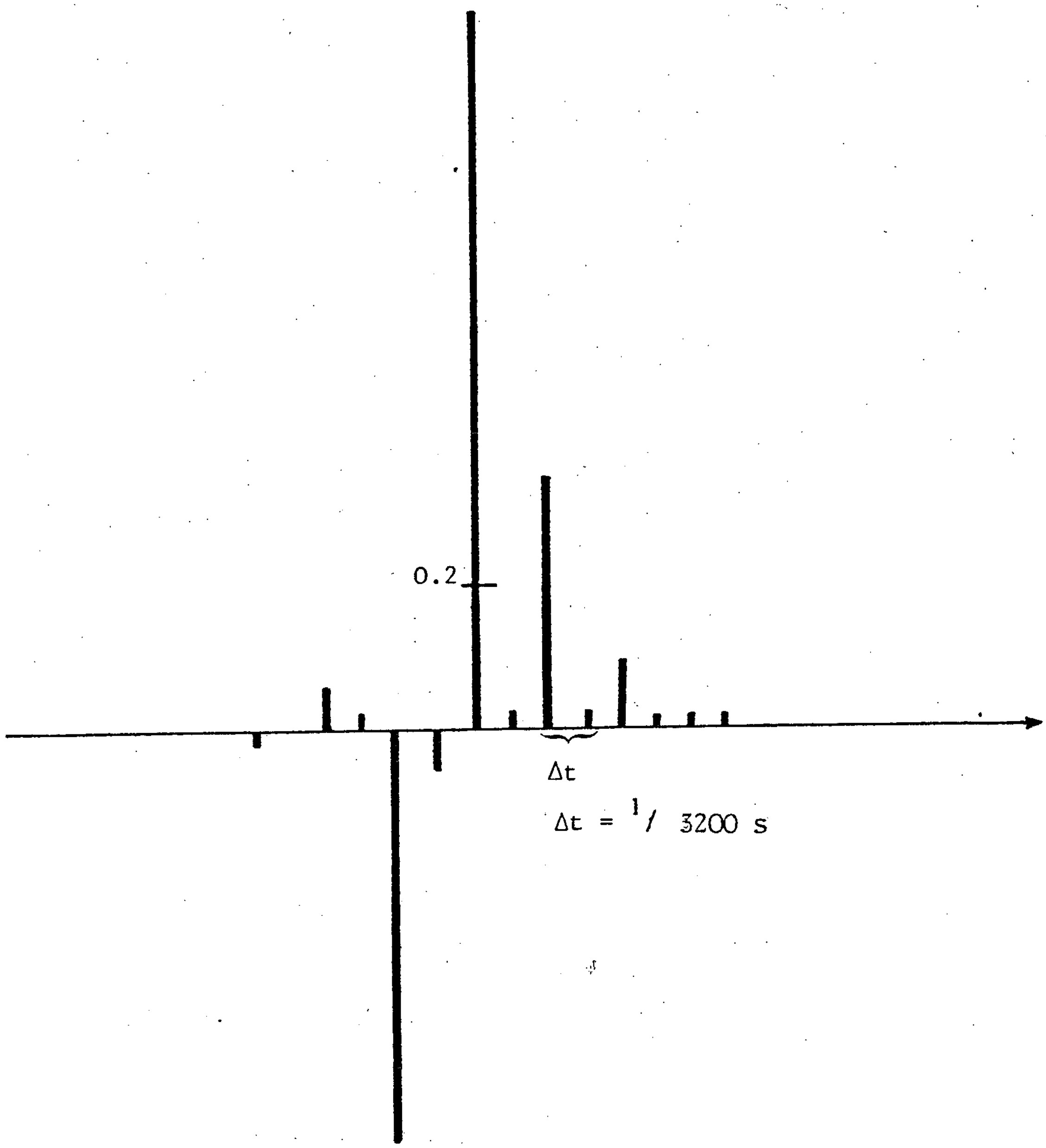



Fig.10 - Sampled impulse response

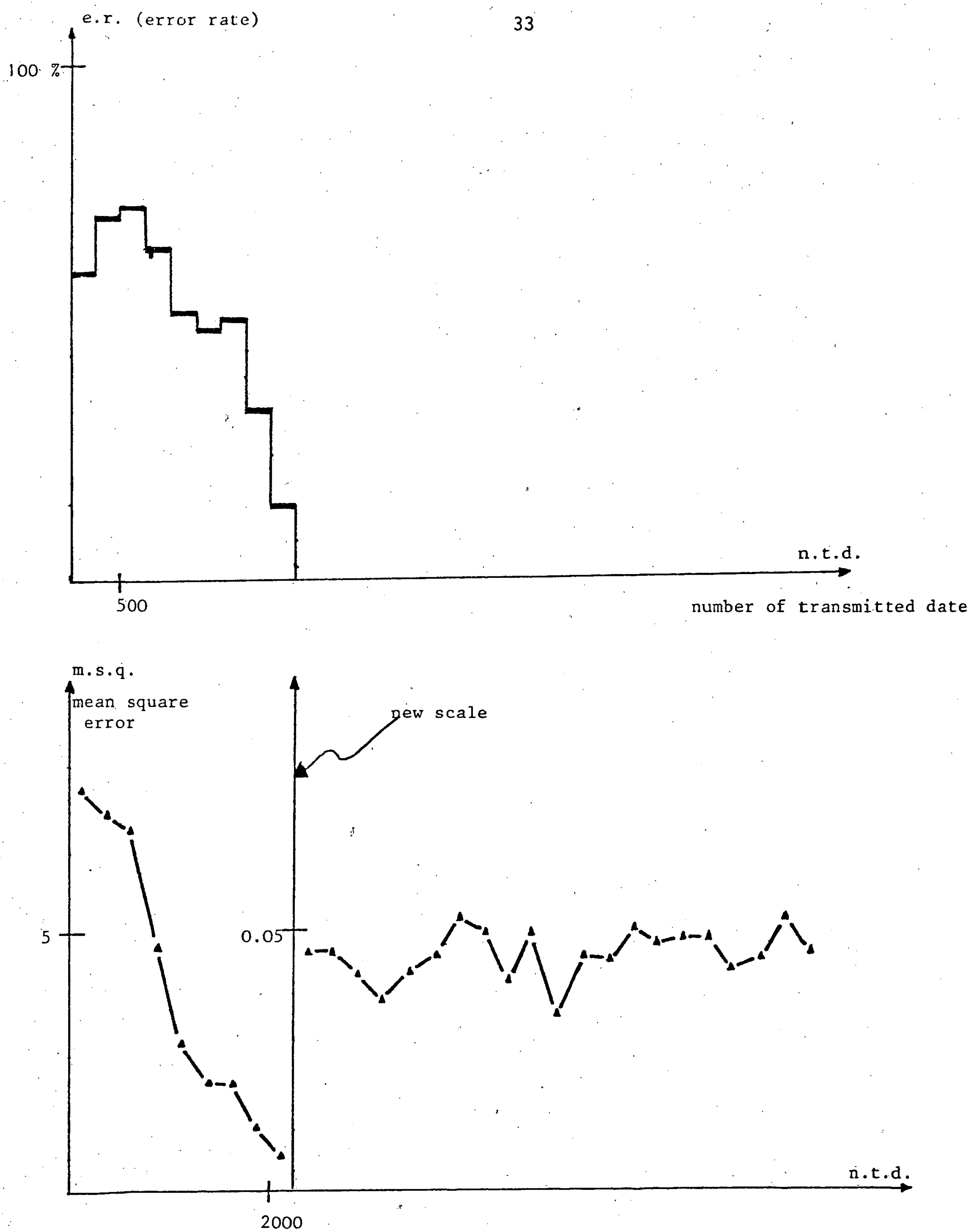



Fig. 11 - Error rate and mean square error for AN1

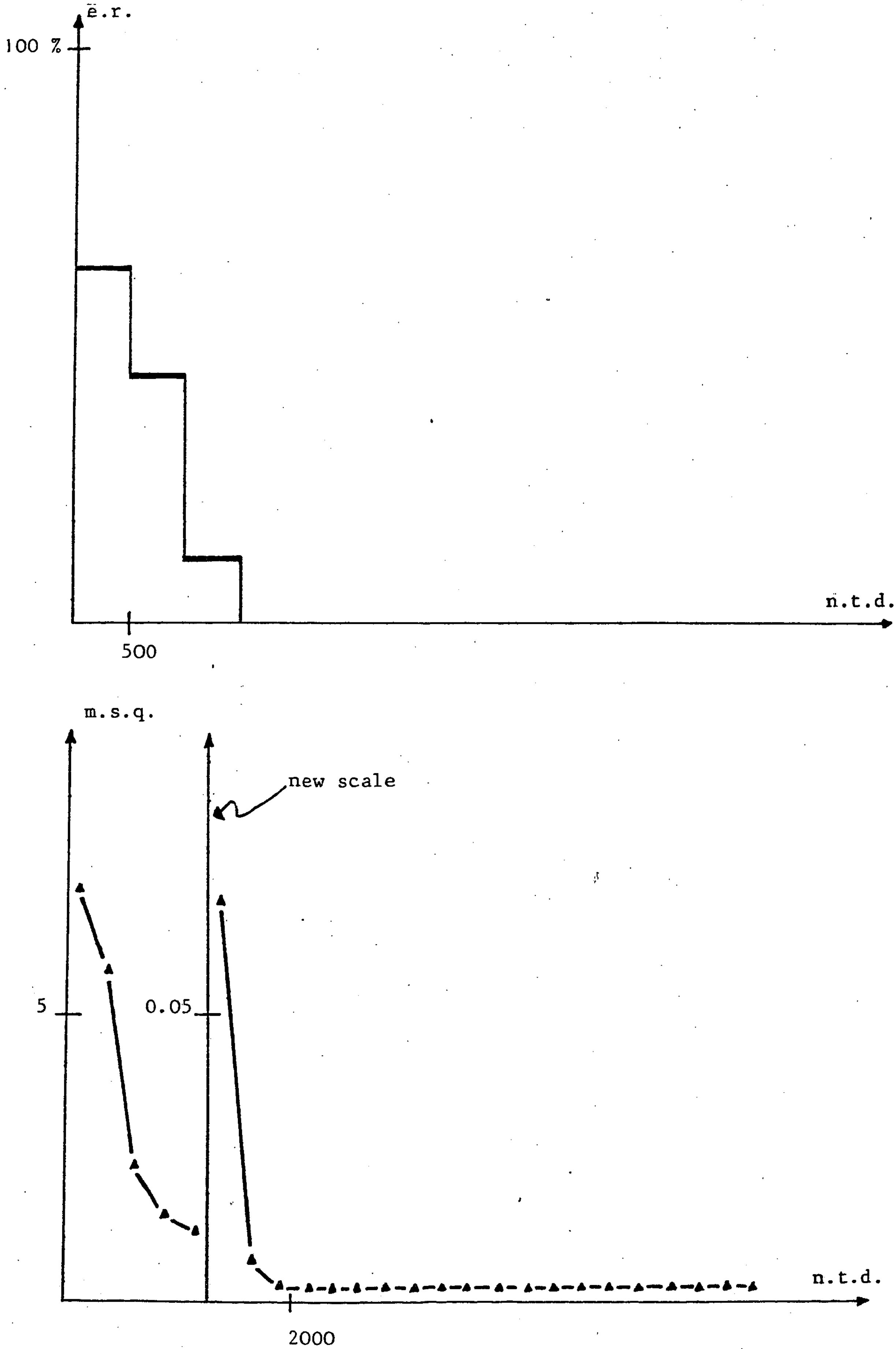



Fig.12 - Error rate and m.s.q. for ASS1

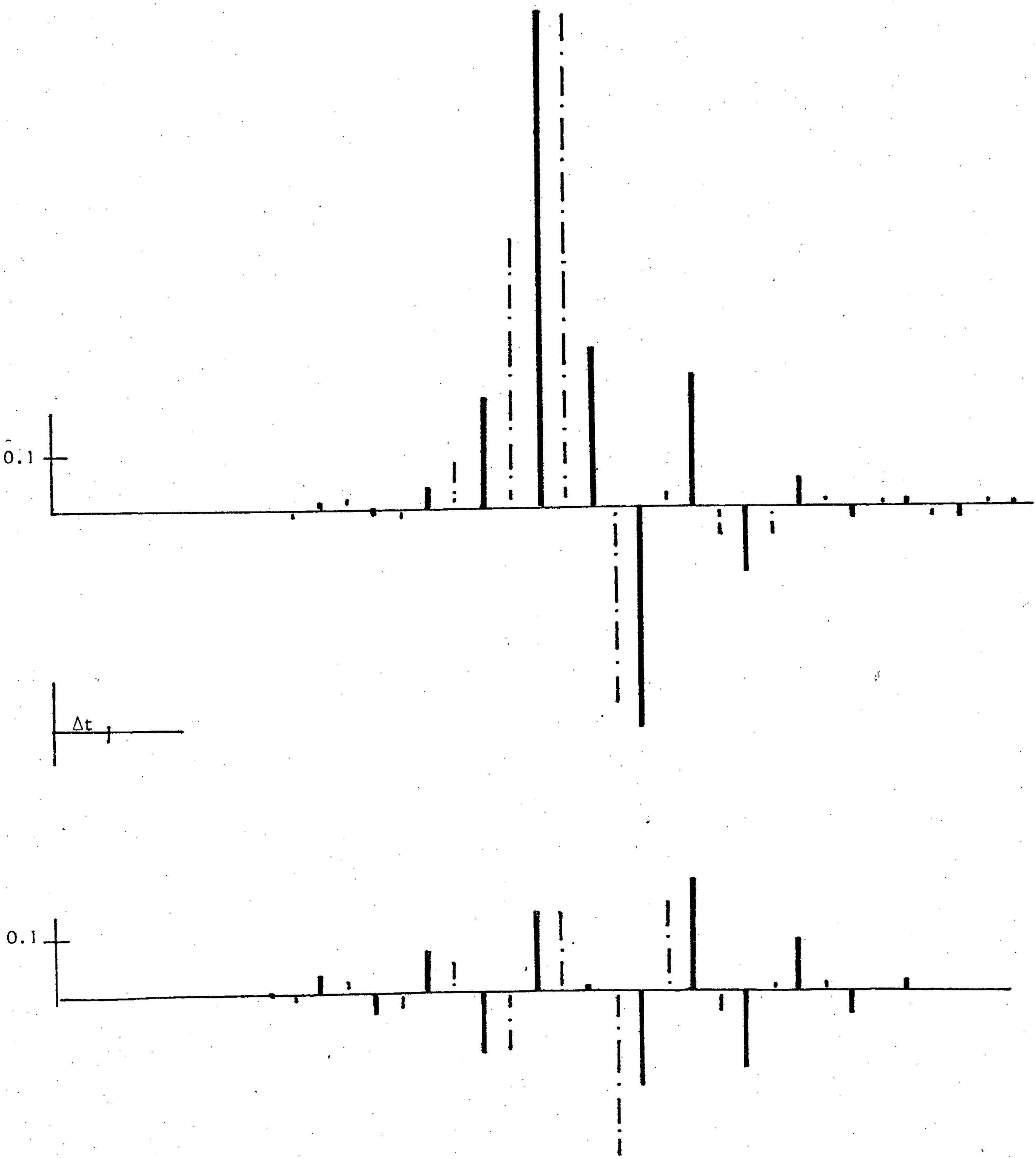



Fig. 13 - Sampled impulse responses

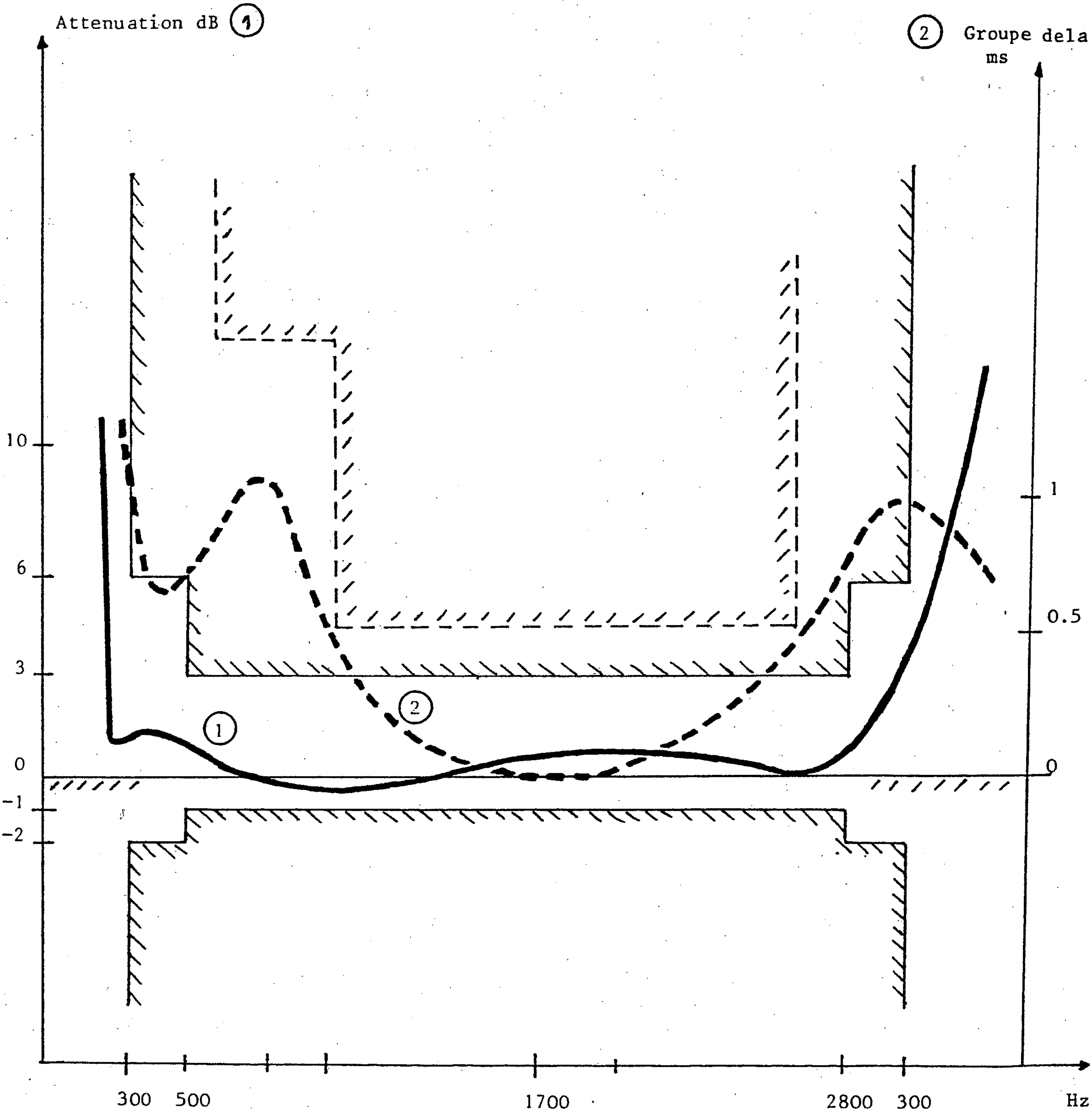
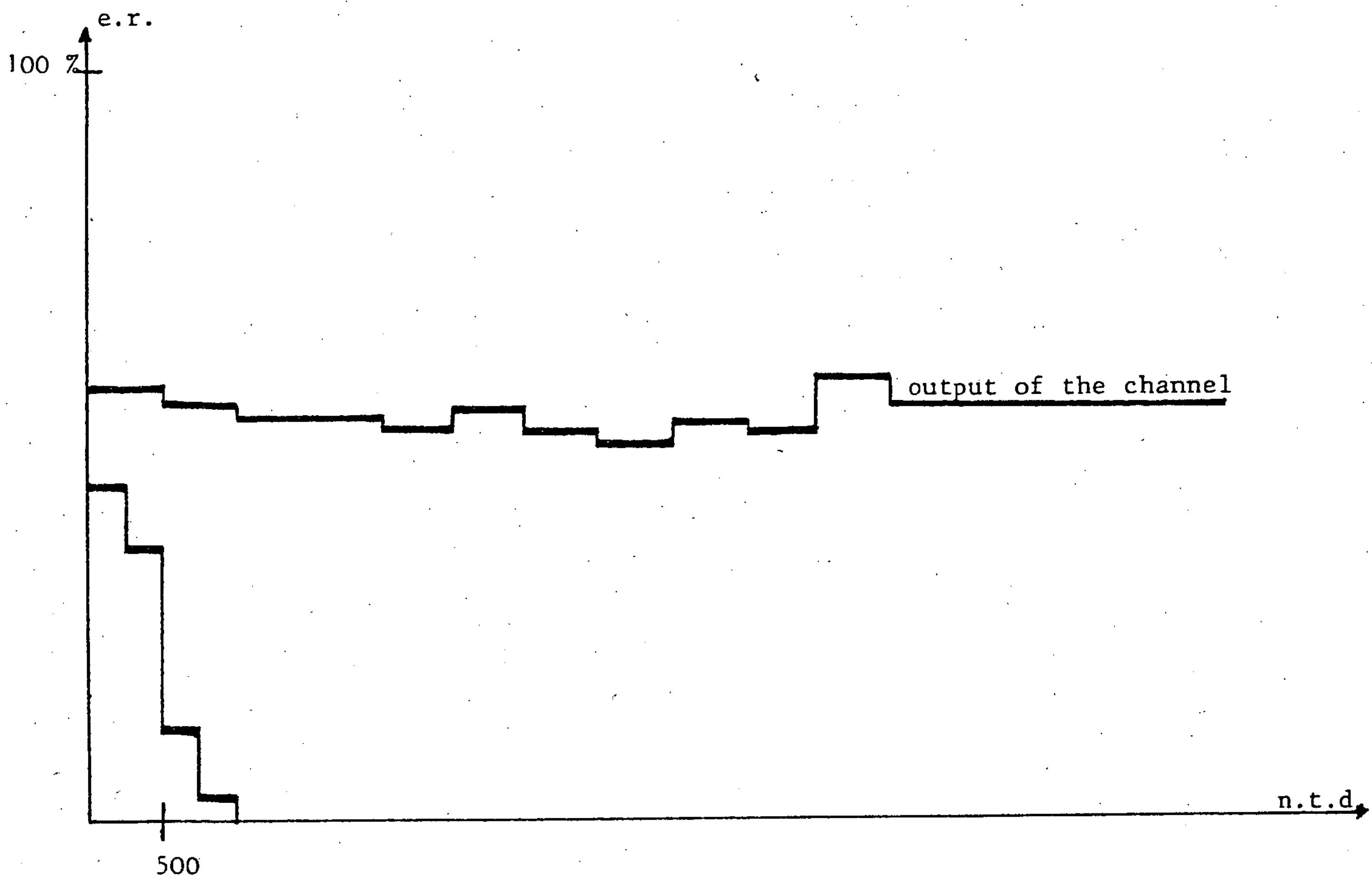



Fig. 14 - Amplitude and delay distorsions of the channel given by fig. 8.



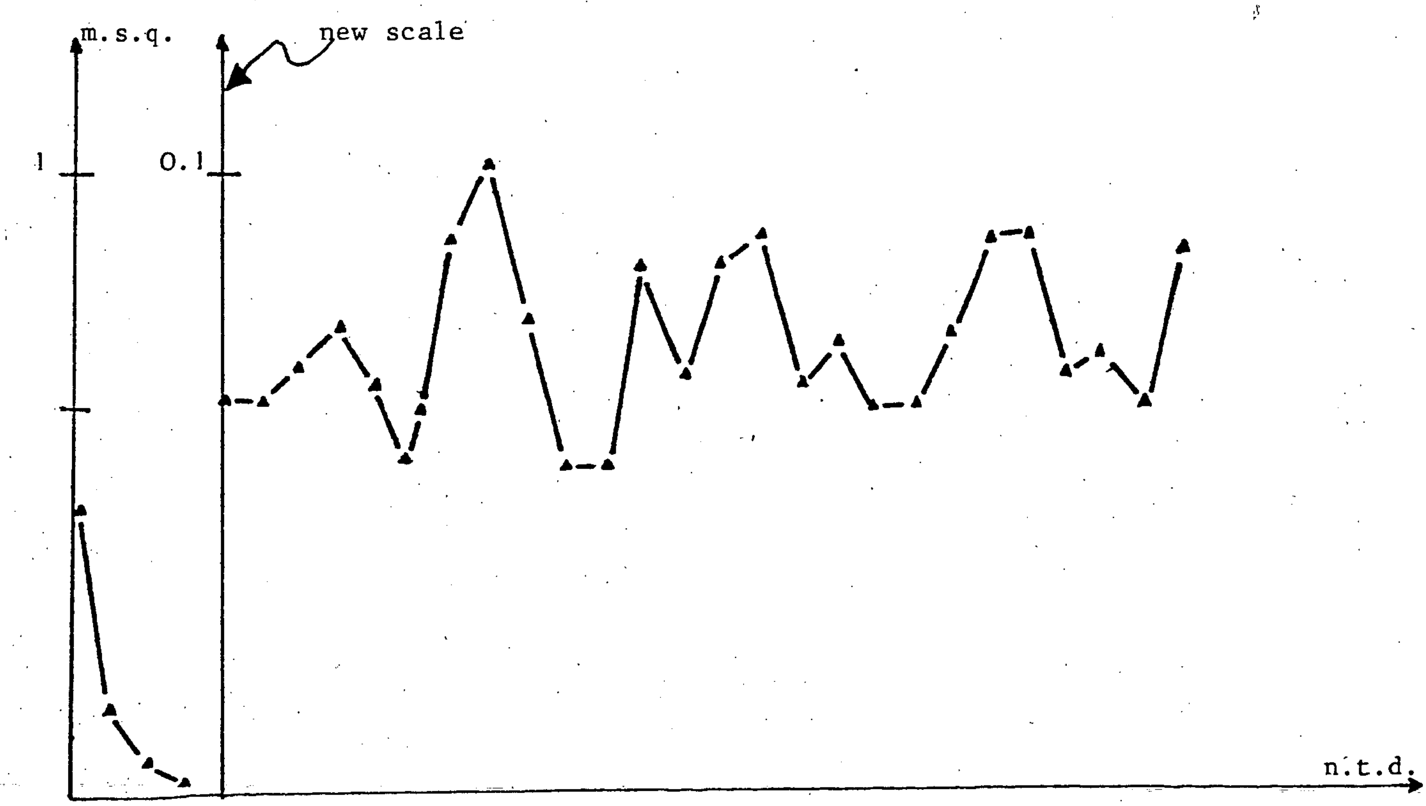




Fig. 15 - Results for AN2 (MAQ transmission)

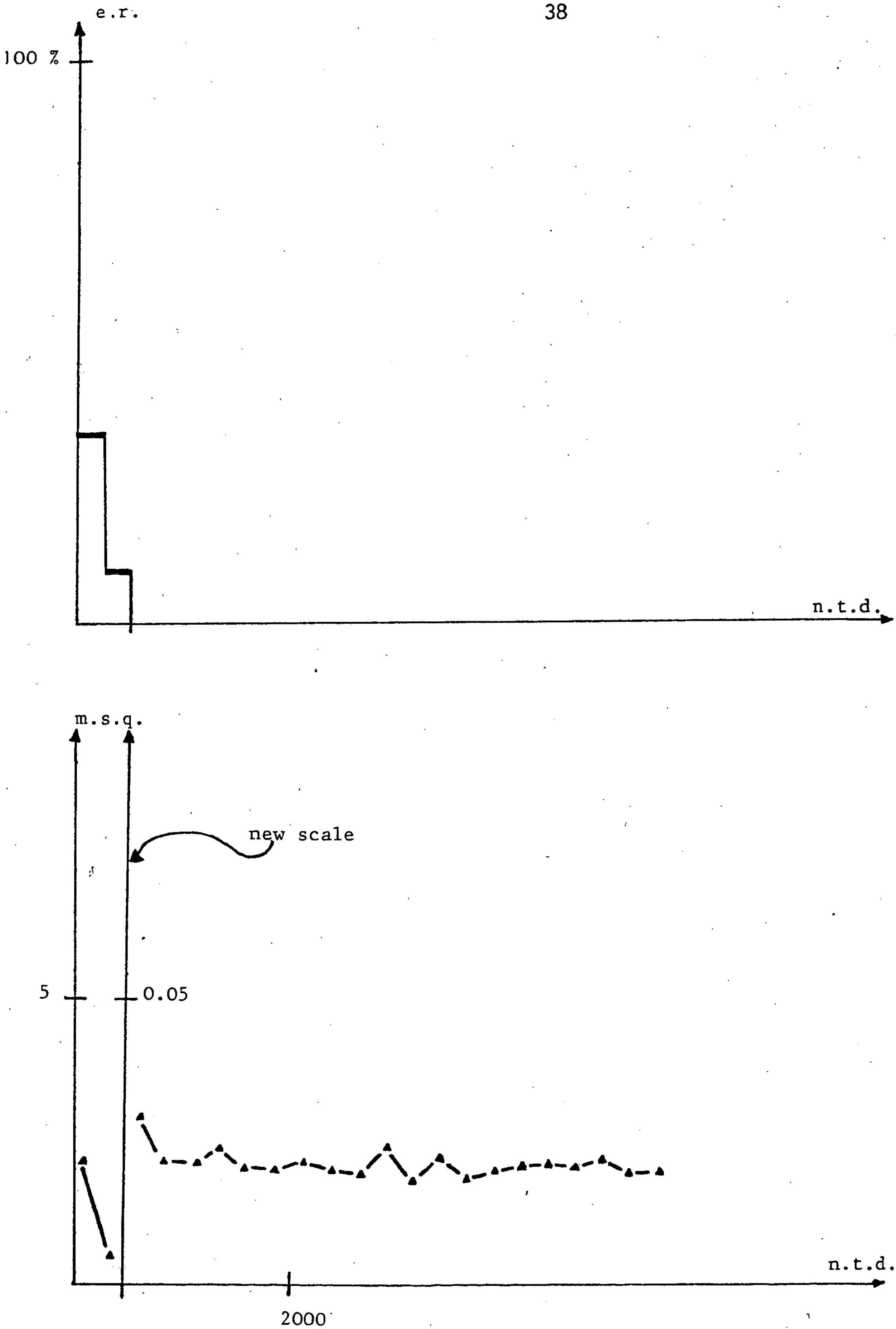
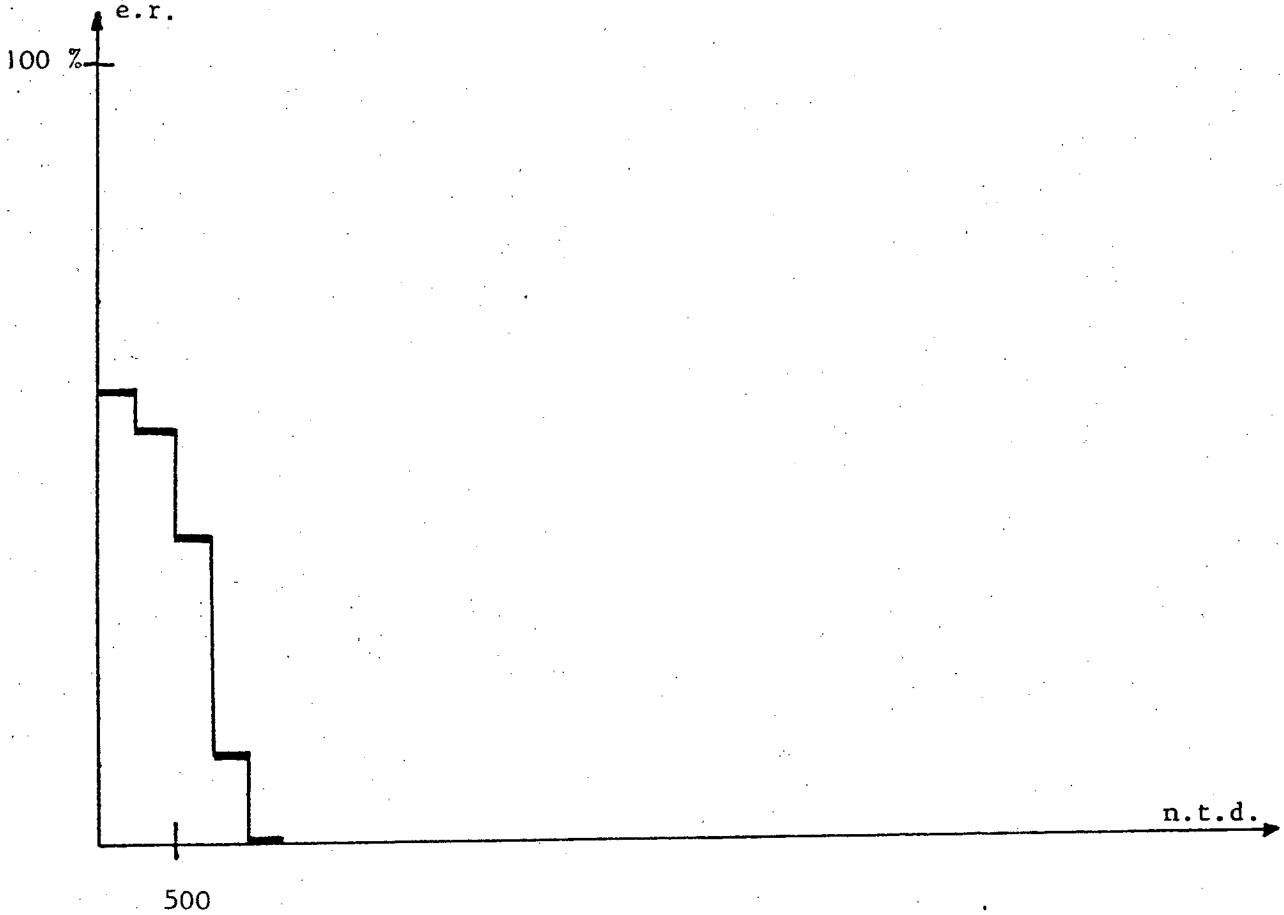



Fig. 16 - Results for ASS2 (MAQ transmission)



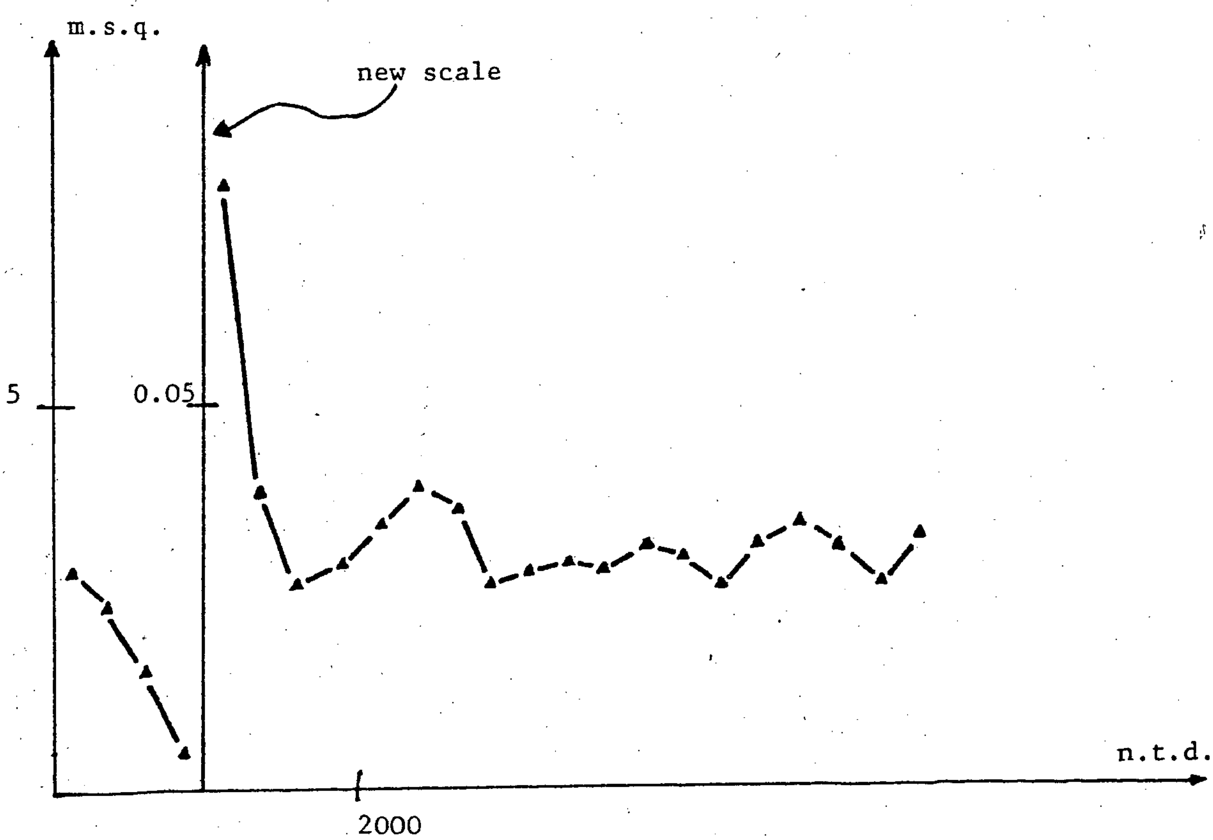




Fig. 17 - Results of ASS2 for the 16 points V 29 constellation.

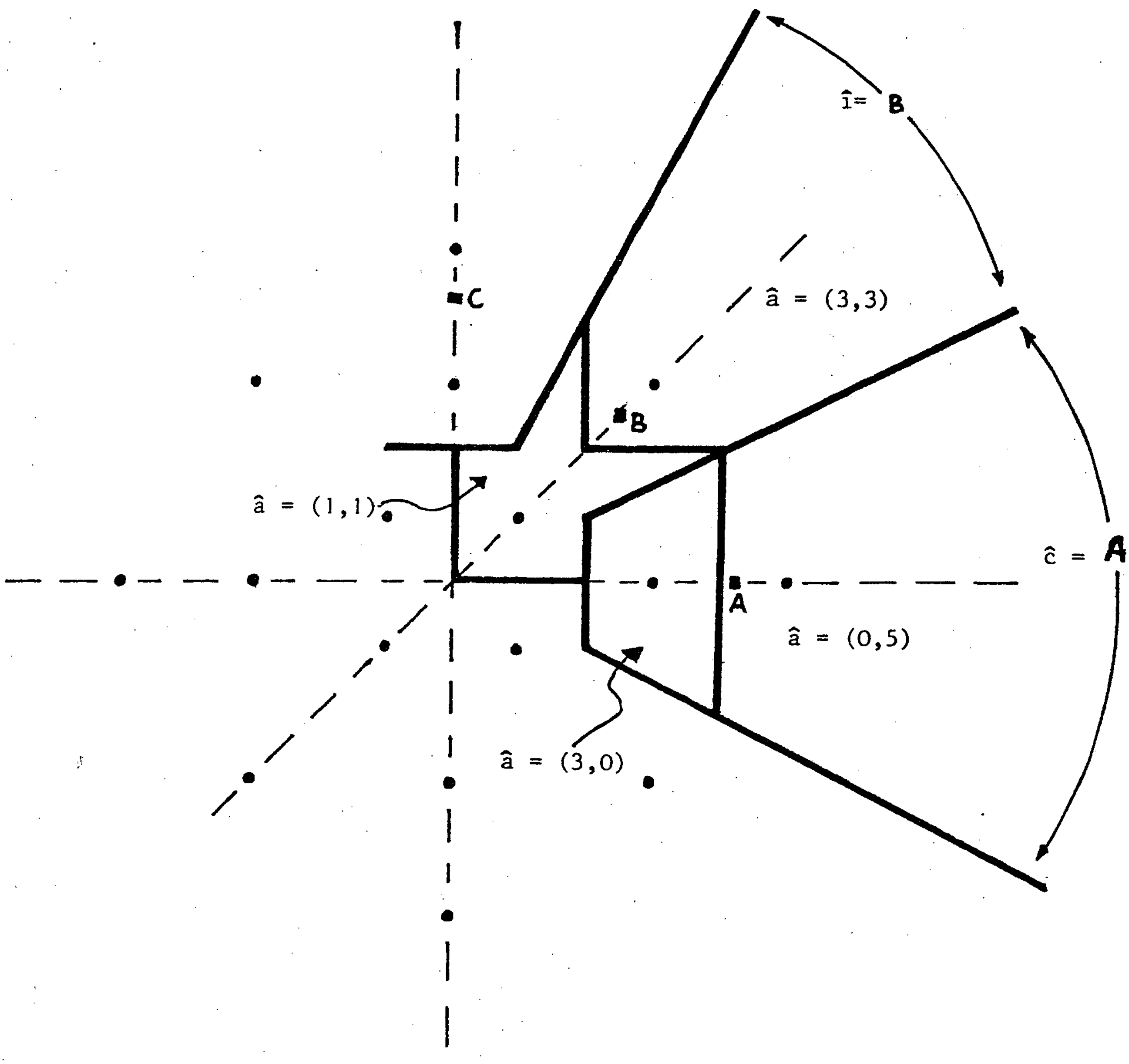
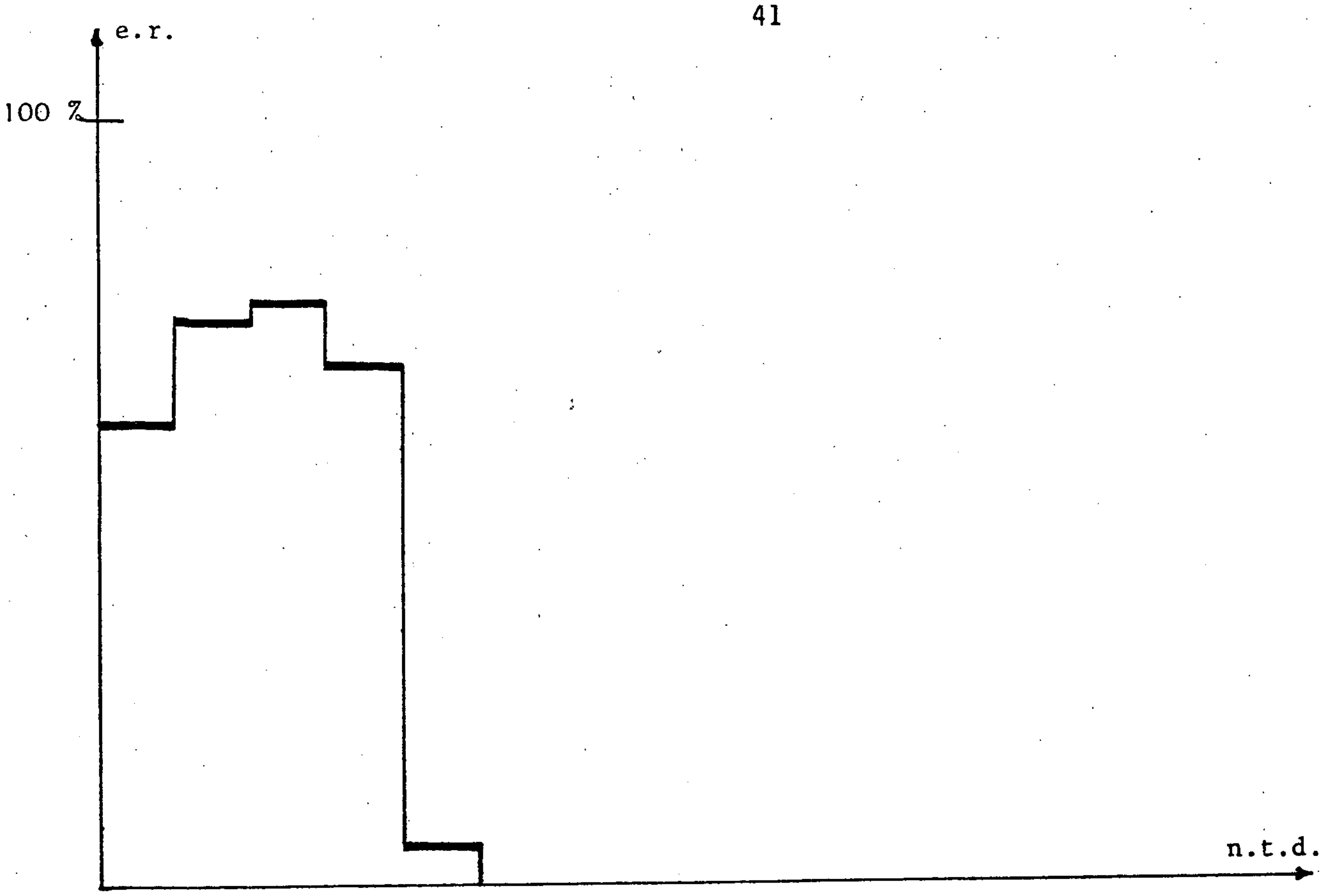



Fig. 18 - Decision areas for the V 29 CCITT constellation



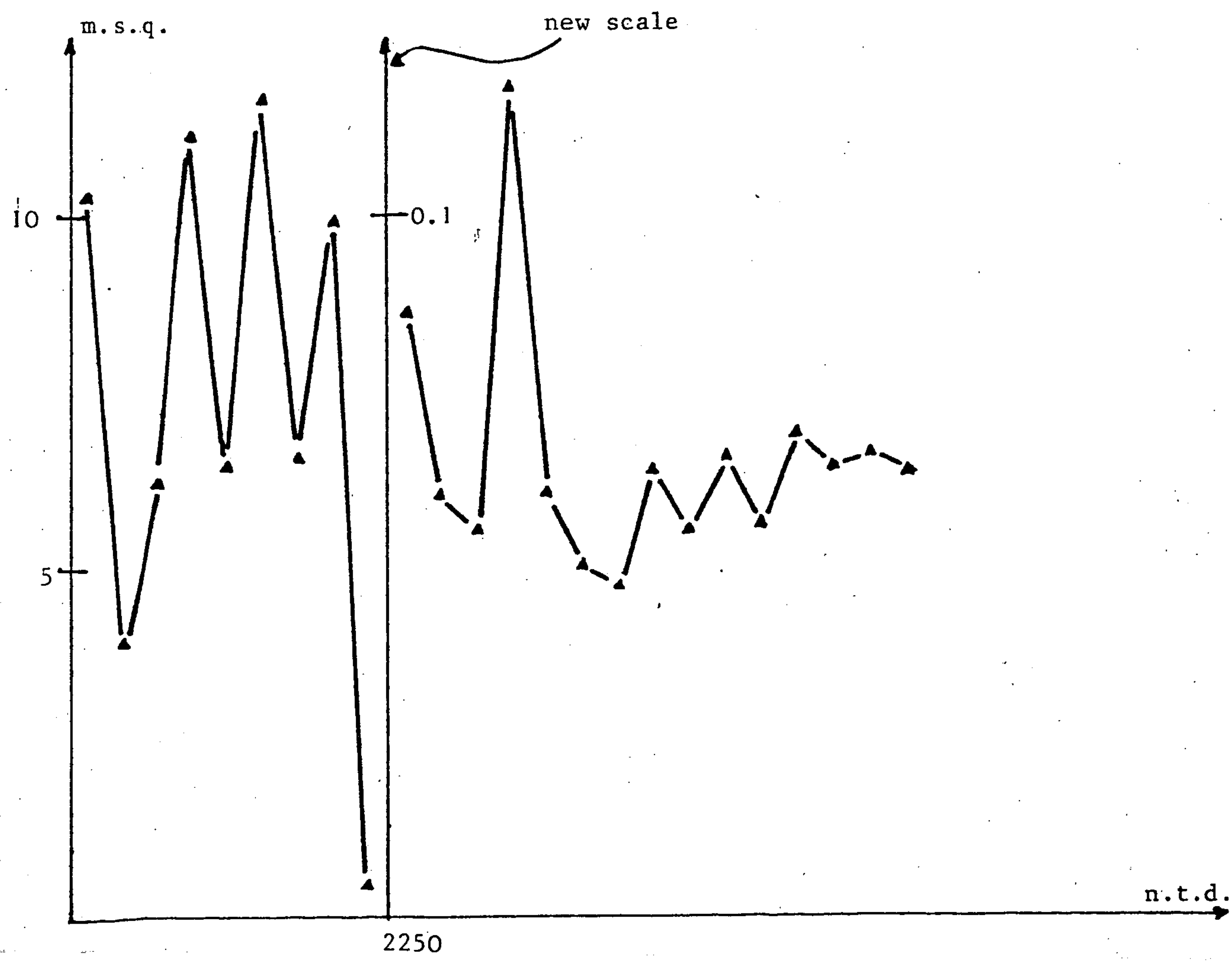




Fig. 19 - Results of ASS2 with a 7 Hz frequency offset

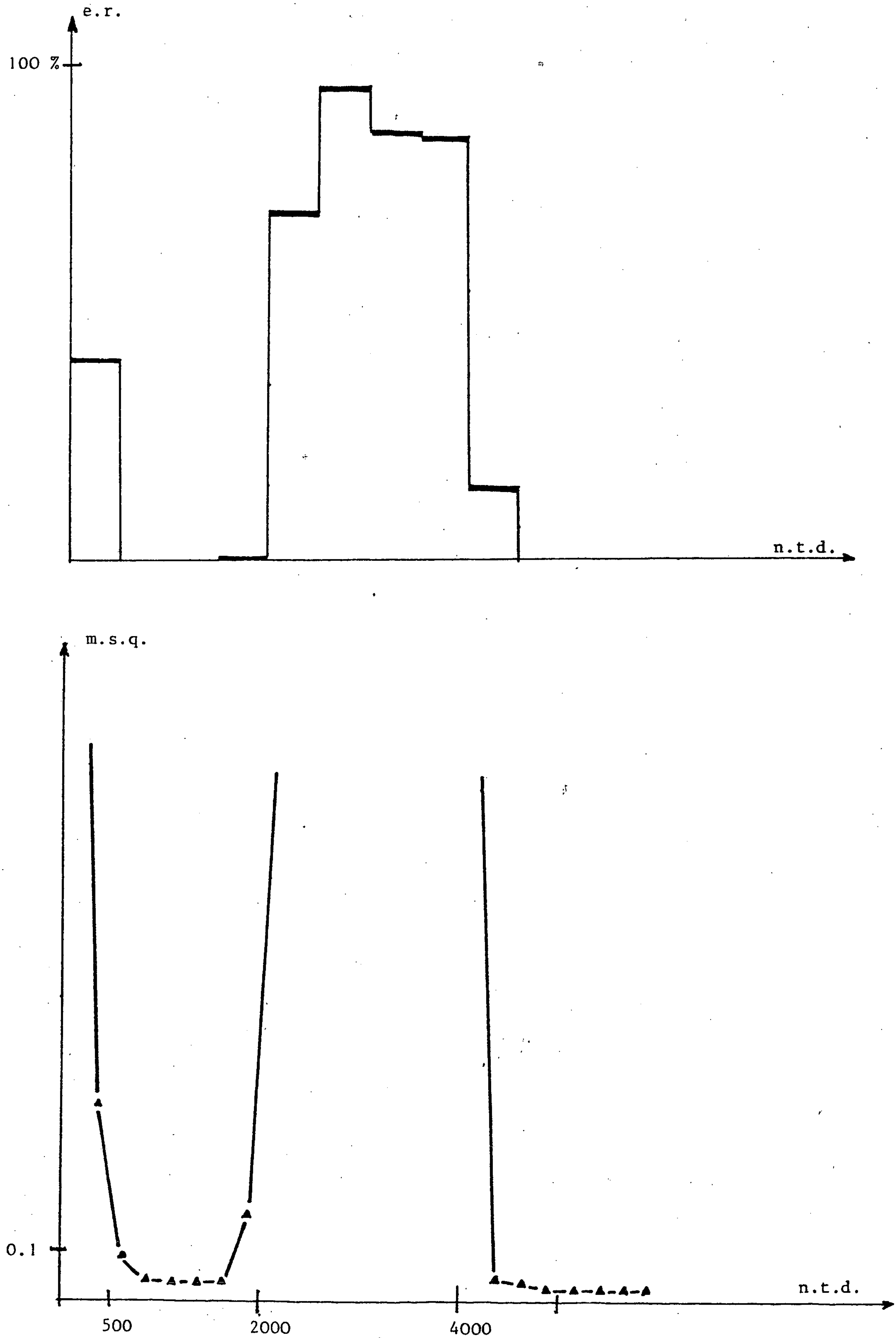



Fig. 20 - Results of ASS2 (2 Hz offset + jump of the channel).

REFERENCES

(1) R.W. Lucky, A survey of the communication theory litterature, 1968-1973. IEEE Trans. on I.T. vol.19, pp 725-739, Nov. 1973.

(2) A.J. Viterbi editor & al., Advances in communications systems vol 4 Academic Press 1975.

(3) R.W. Lucky, J. Salz, E.J. Weldon, Principles of data communication, Mc Graw Hill, 1968.

(4) R.D. Gitlin, F.R. Magee, Self orthogonalizing adaptive equalization algorithms, IEEE Com. vol.25, n°7, pp 666-672, July 1977.

(5) D.D. Falconer, Jointly adaptive equalization and carrier recorvery in two-dimensionnal digital communication systems, Bell S.T.J. vol.55, pp 316-334, 1976.

(6) Y. Sato, A method of self-recovering equalization for multilevel amplitude-modulation systems, IEEE Trans. on Com. vol 23, pp 679-682 June 75.

(7) D.N. Godard, Self recovering equalization and carrier tracking in two-dimensional data communication systems, IEEE Trans. on Com. vol 28, pp 1867-1875, Nov. 1980

(8) A. Benveniste, M. Bonnet, M. Goursat, C. Macchi, G. Ruget, INRIA report 325, 1978.

(9) A. Benveniste, M. Goursat, G. Ruget, Robust identification of a non minimum phase system : blind adjustment of a linear equalizer in data communication, IEEE Trans. on A.C. vol 25, pp 385-399, June 1980.

(10) A. Benveniste, M. Goursat, G. Ruget, Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithms, IEEE Trans. on A.C. vol 25, pp 1042-1058, Dec. 1980.

(11) Contract INRIA/CNET, Report 1982.

(12) A. Benveniste, G. Ruget, A measure of the tracking capability of recursive stochastic algorithms with constant gains, IEEE Trans. A.C. vol 27, June 1982.

(13) A. Benveniste, M. Goursat, A gain and phase identification procedure : blind adjsutment of a recursive equalizer, IEEE Inf. Th. Symposium Grignano, June 1980.

(14) L. Ljung, On positive real transfer functions ... IEEE A.C. 22, 1977.

(15) D. Falconer, Analysis of a gradient algorithm for simultaneous passband equalization and carrier phase recovery, BSTJ, April 1976, pp. 409-428.

(16) D. Falconer & J. Salz, Optimal reception of digital data over the gaussian channel with unknown delay and phase jitter, IEEE I.T. vol 23, Jan. 1977.

(17) L. Ljung, Analysis of recursive stochastic algorithms, IEEE A.C. 22, 1977.

(18) T. Soderstrom, Ljung, I. Gustavsson, A theoretical analysis of recursive identification methods, Automatica vol 14, 1978.

## 2. A GAIN AND PHASE IDENTIFICATION PROCEDURE : BLIND ADJUSTMENT OF A RECURSIVE EQUALIZER.

# A GAIN AND PHASE IDENTIFICATION PROCEDURE : BLIND ADJUSTMENT OF A RECURSIVE EQUALIZER.

Consider an unknown linear time-invariant system S without control, drived by random variables with known law. We are interested in the identification of S from the output. The usual results work only under the major assumption : S is minimum phase. In our case the system S is non minimum phase and the litterature gives only a negative result : the identification is impossible for a gaussian driving noise. For a large class of other input laws we present an identification procedure in ARMA form and present some numerical results for a concrete case, origin of our study : the blind settling phase of a recursive equalizer in data communication.

## THE PHYSICAL PROBLEM

The problem we describe here is related to the data transmission with a telephone line. The transmission system may be considered as an unknown time-invariant linear system S without control. The input is a sequence of independent identically distributed (i.i.d.) random variables (r.v.) which are the data to be transmitted. The duration of the impulse response of S being large with respect to each data, we have intersymbol interferences in the output. The problem is to restore the unknown transmitted sequence or, equivalently, identify the inverse channel $S^{-1}$ from the observations. The solution consists in placing a serial adaptive system called equalizer which is a linear filter ; its tap weights are adjusted in order to minimize the mean square error for its output. The adjustment is done by a stochastic gradient algorithm. This technique requires knowledge of transmitted data : this difficulty is avoided in practice by a settling phase during which the emitter transmits an à priori known sequence whith which we get an equalizer close to the inverse. The obtained tap weights give good estimates of the input ; the algorithm continues by using the estimates and the equalizer follows the evolution of the inverse (in practice, obviously, the channel is not time-invariant but the adaptivity of the algorithm is quick enough for the non-stationarity of the channel). In some cases (for example : break in a multipoint communication) the transmission of a known sequence is impossible and the receiver is obliged to inverse the system S only with the receiver output flow : this is our blind adjustment problem. In [1], [3], we have given algorithms for adjusting transversal equalizers. Here we give algorithms for adjusting recursive equalizers : new problems appear like instability, error propagation.

Remark : We do here a little mistake because we neglect the additive noise on the output. The previous procedure (minimizing the mean square error) realizes simultaneously this work. We will see that neglecting the noise is not a worry in our case.

## §1 : FORMULATION OF THE PROBLEM.

Let $(x_t)_{t\in Z}$, $(a_t)$ being an unobserved i.i.d. random sequence with known distribution $\nu$. How can one identify S, observing only the output $(x_t)$ ?

Suppose we know that S is a stable minimum phase system (i.e. with stable inverse) ; then the problem is very classical and $(a_t)$ is simply the innovation of $(x_t)$, defined by

(1) $\quad a_t = \lambda(x_t - \hat{x}_t)$,

$\hat{x}_t$ being the best linear prediction of $x_t$ given $x_{t-1}, x_{t-2}, \ldots$, and $\lambda$ a constant normalizing the variance. Various algorithms in both AR and ARMA form are wellknown for such an identification of S. ([5], [6]).

The difficult point in our problem is that S cannot be assumed to be minimum phase ; so we have to identify both gain and phase of S, the later being impossible with the aid of the second order statistics only (and even impossible if $(a_t)$ is a gaussian white noise because second order statistics of $(x_t)$ are exhaustive in that case).

Throughout the whole paper, the following assumption is in force :

A1 : *(i)* $\nu$ is symmetric with finite variance ;

*(ii)* S and $S^{-1}$ (the inverse of S) are of finite energy, i.e. the covariance matrix

$$\Lambda = (\Lambda_{i,j})_{i,j \in Z}, \quad \Lambda_{ij} = \mathbb{E}(x_i x_j)$$

is a bounded and positive definite operator on the Hilbert space $\ell^2$ of square convergent series.

REMARKS

1/ S non minimum phase, thus $S^{-1}$ cannot be a causal system ; so we cannot restore $(a_t)$ on-line, but only the whole sequence off-line ; in practice we shall have restoration on-line with a delay. Anyway, $S^{-1}$ is defined up to a time shift.

2/ $\nu$ being symmetric, both $(x_t)$ and $(-x_t)$ have the same joint distribution, so that we cannot distinguish between S and -S.

Therefore the best we can achieve is (see *fig. 1*).

Goal G : Find a (possibly non-causal) system H so that the global system $T = H \circ S$ is $\pm$ identity up to a time shift (i.e. T is, up to a sign change, a pure delay).

## §2 : A GAIN-AND-PHASE IDENTIFYING FUNCTIONAL.

For further discussion and explanation, see [1], [2], [3]. From now on $\nu$ is necessary non-Gaussian ; furthermore, let us assume that $\nu$ is subgaussian in the following sense :

A2 : *(i)* or *(ii)* is fullfilled :

*(i)* $\nu$ is uniform over $[-\ell,+\ell]$ ;

*(ii)* $\nu(dx) = k\, e^{-g(x)}dx$, g even, g(x) and $\frac{g'(x)}{x}$ strictly increasing over $\mathbb{R}_+$.

Example : $\nu(dx) = k\, e^{-|x|^\gamma}dx, \quad \gamma > 2$

Define the following functional $\mathcal{J}(H)$ (H adjustable system) by (see *fig. 1*).

$$(2) \qquad \mathcal{J}(H) = \mathbb{E}(c_t^{\,2} - 2\alpha|c_t|), \quad \alpha = \frac{\mathbb{E}\, a_t^{\,2}}{\mathbb{E}|a_t|},$$

$(c_t)$ output of the adjustable system H with $(x_t)$ as input. Then we have the following result : ([1], [2]).

THEOREM 1 Under A1 and A2, the only local (and hence global) minima of $\mathcal{J}$ are the systems $\pm S^{-1}$, (up to a time shift).

So, by theorem 1, minimization of the functional $\mathcal{J}$ achieves goal G.

## §3 : RECURSIVE GRADIENT AND EXTENDED GRADIENT ALGORITHMS FOR MINIMIZING $\mathcal{J}$ WITH AN ARMA FORM FOR H.

In [1], we have used a moving average representation of S in A.R. form. Here we want identification in ARMA form. Using the transfer function notation, let for instance $S(z^{-1})$ be of the following form

$$(3) \qquad S(z^{-1}) = \frac{g\, P_1(z^{-1})\, P_2(z^{-1})}{Q(z^{-1})},$$

with g a gain, Q and $P_1$ stable monic polynomials (i.e. with zeros outside the unit circle) and $P_2$ unstable monic polynomial (with zeros inside the unit circle) ; since S is non-minimum phase, $P_2$ cannot be a constant. We expand $\frac{1}{P_2(z^{-1})}$ in a Laurent serie

$$(4) \qquad \frac{1}{P_2(z^{-1})} = \sum_{k \geq 0} \alpha_k\, z^k$$

which converges outside the unit circle. Then we have

$$(5) \qquad \frac{1}{S(z^{-1})} = \frac{1}{g}\; \frac{Q(z^{-1})\, \left(\sum_{k \geq 0} \alpha_k z^k\right)}{P_1(z^{-1})},$$

and, finally, truncating the Laurent Serie, we obtain the following stable but non-causal (with delay n) approximation of $\frac{1}{S}$ :

(6) $$\frac{1}{S(z^{-1})} = \frac{1}{g}\,\frac{Q(z^{-1})\,\left(\sum_{k=0}^{n} \alpha_k z^k\right)}{P_1(z^{-1})},$$

which enables us to choose the following form for the adjustable system H

(7) $$c_t = k_1\, c_{t-1} + \dots + k_p\, c_{t-p} + h_{-n} x_{t+n} + \dots + h_0\, x_t + \dots + h_n x_{t-n},$$

or, in transfer function form

(8) $$c_t = \frac{H(z^{-1})}{1-K(z^{-1})}\, x_t$$
$$H(z^{-1}) = \sum_{-n}^{n} h_i Z^{-i}, \quad K(z^{-1}) = \sum_{1}^{p} k_i z^{-i},$$

the zero of 1-K being outside the unit circle.

Notation : If $(y_t)$ is a stationary sequence, let us denote by $(y_t{}^K)$ the filtered sequence defined by

(9) $$y_t{}^K = \frac{1}{1 - K(z^{-1})}\, y_t.$$

Let

(10) $$(*)\,{}^t\theta = (k_1,\dots,k_p\,;\, h_{-n},\dots,h_0,\dots,h_n)$$
$${}^t\phi_t = (c_{t-1},\dots,c_{t-p}\,; x_{t+n},\dots,x_t,\dots,x_{t-n})$$

be respectively the vector of adjustable parameters and preceding observations (with delay n). Then a straight-forward calculation gives

(11) $$\frac{d}{d\theta}\mathcal{J}(\theta) = \mathbb{E}\,(\phi_t{}^K(c_t - \alpha\ \mathrm{sgn}\ c_t),$$

$c_t$ output of $\frac{H(z^{-1})}{1-K(z^{-1})}$ with adjustable coefficients given in the vector $\theta$ and sequence $(\phi_t{}^K)$ obtained by filtering $(\phi_t)$ through $\frac{1}{1-K(z^{-1})}$ .

A recursive stochastic gradient algorithm (RSGA).

In the same way as one obtains a stochastic gradient form for the R.M.L. algorithm ([6]) in the case of a minimum-phase identification, we derive from (7) the following stochastic gradient procedure

(12) $$\begin{cases} \theta(t+1) = \theta(t) - \tau\, \phi_t{}^K(c_t - \alpha\,\mathrm{sgn}\, c_t)\ (\tau>0 \text{ small}) \\ c_t \quad = {}^t\theta(t)\ \phi_t \\ c_t{}^K \quad = k_1(t)\ c^K_{t-1} + \dots + k_p(t)\ c^K_{t-p} + c_t \\ x^K_{t+n} = k_1(t)\ x^K_{t+n-1} + \dots + k_p(t)\ x^K_{t+n-p} + x_{t+n} \end{cases}$$

Remark that the sequences $(c_t)$, $(c_t{}^K)$, $(x_t{}^K)$ are only approximate of $\frac{H(z^{-1})(t)}{1-K(z^{-1})(t)}\,x_t$, etc... .

For the justification of obtaining (12) from (11), see for instance [4], [7].

An extended recursive stochastic gradient algorithm (ERSGA).

Following the analogy with the (classical) minimum phase case, we obtain the extended-least-squares-type ([6]) version of the gradient algorithm, without filtering $(\phi_t)$ :

(13) $$\begin{cases} \theta(t+1) = \theta(t) - \tau\ \phi_t\ (c_t - \alpha\ \mathrm{sgn} c_t) \\ c_t \quad = {}^t\theta(t)\ \phi_t. \end{cases}$$

REMARK 3 :

Algorithm (13) looks like a classical recursive equalizer in which the estimated error signal $(c_t - \alpha\ \mathrm{sgn} c_t)$ is modified ; in fact, if the input takes only two values ± 1, (13) is the classical recursive equalizer, which is hence seen to be robust in this case. Filtering the state in RSGA prevents in a certain sense from error propagation.

## §4 : THEORETICAL ANALYSIS OF RSGA AND ERSGA.

Convexity of the functional $\mathcal{J}$

Concerning RSGA, general results on stochastic approximation ([4] or [7]) ensures convergence as long as $1-K(z^{-1})$ remains a stable polynomial (a condition that can be enforced). As an interpretation of the efficiency of RSGA near the optimum, we shall calculate $\frac{d^2}{d\theta^2}\mathcal{J}(\theta^*)$ under the assumption that $\frac{H(^*z^{-1})}{1-K^*(z^{-1})}$ is the minimum order representation of $S^{-1}$, up to a time shift (See Remark 4 below).

(*) *The superscript* $^t$. *denotes transposition*

Let $\nu(dx) = f(x)dx$ be sufficiently smooth for $\frac{d^2}{d\theta^2} J$ to exist. Then we have (see Appendix 1) :

$$\frac{d^2}{d\theta^2} J(\theta^*) = (1-2\alpha f(0))\, \mathbb{E}\left[\phi_t^{K^*}\ {}^t\phi_t^{K^*}\right] \tag{14}$$
$$+ 2\alpha f(0)\ \mathbb{E}\left[(\hat{\phi}_t^{K^*}|a_t)\ {}^t(\hat{\phi}_t^{K^*}|a_t)\right]^{(+),(++)}.$$

It can be proved that, under A2, $1-2\alpha f(0) > 0$ (it is $=0$ for $\nu$ Gaussian) ; on the other hand, the covariance matrix $\mathbb{E}(\phi_t^{K^*}\ {}^t\phi_t^{K^*})$ is positive definite because of minimality of the representation $\frac{H^*}{1-K^*}$ of $S^{-1}$ ; the second covariance matrix in (10) is of rank 1.

Compare (10) with the second derivative of the least squares functional used in minimum-phase ARMA identification, where it is equal to $\mathbb{E}\left[\phi_t^{K^*}\ {}^t\phi_t^{K^*}\right]$ ; in the most favorable case of uniform distribution, we get

$$1 - 2\alpha f(0) = \frac{1}{3},$$

which relates the efficiency of our algorithm to the least squares one (which can be used only in minimum phase situations). In the Gaussian case, $\frac{d^2}{d\theta^2} J$ becomes singular, which is not surprising ; it remains only the second term in (10) that signifies good adjustment of the variance of the output $(c_t)$ of the adjustable system.

REMARK 4 :

The assumption that $\frac{H^*}{1-K^*}$ is the "exact minimal order representation" for $S^{-1}$ cannot obviously be valid (except for the minimum phase case) because formula (6) says clearely that any finite order representation is an approximation of $S^{-1}$. This theoretical point will be further analyzed elsewhere.

(+) *Here* $\phi_t^{K^*}$ denotes $(\phi_t^*)^{K^*}$, *where* $\phi_t^*$ is built *with the output of* $\frac{H^*(z^{-1})}{1-K^*(z^{-1})}$ *with input* $(x_t)$.

(++) $(\hat{\phi}_t^{K^*}|a_t)$ *denotes the best linear estimate of* $\phi_t^{K^*}$ *given* $a_t$.

Local analysis of ERSGA

It is now wellknown ([4], [7]) that the asymptotic behaviour of (13) as $\tau \searrow 0$ (as long as the system remains stable) is given by the following differential equation

$$\begin{cases} \frac{d}{ds}\theta = -V(\theta) \\ V(\theta) = \mathbb{E}(\phi_t(c_t - \alpha\, \mathrm{sgn}\, c_t)), \quad c_t = {}^t\theta\, \phi_t. \end{cases} \tag{15}$$

For example, local convergence of (13) to $\theta^*$ is ensured if and only if $\theta^*$ is a stable equilibrium of (15), whereas global convergence requires global stability of (15).

Let us analyze local convergence : first of all

$$0 = \frac{d}{d\theta} J(\theta^*) = \mathbb{E}(\phi_t^{K^*}(c_t - \alpha\, \mathrm{sgn}\, c_t^*))$$

ensures, because $1-K^*$ is stable,

$$0 = \mathbb{E}(\phi_t^*(c_t^* - \alpha\, \mathrm{sgn}\, c_t^*)) = V(\theta^*),$$

so that $\theta^*$ is an equilibrium of (15). For analyzing the stability of this equilibrium, we have, as in (14) :

$$\frac{d}{d\theta} V(\theta^*) = (1 - 2\alpha f(0))\ \mathbb{E}(\phi_t^*\ {}^t\phi_t^{K^*}) \tag{16}$$
$$+ 2\alpha f(0)\ \mathbb{E}((\hat{\phi}_t^*|a_t)(\hat{\phi}_t^{K^*}|a_t)),$$

the difference with (14) lying in the non filtered sequence $\phi_t^*$. Then the condition for local stability of (13) looks like the extended least squares case [6], except for the second term of rank 1. For instance, we have :

THEOREM 2 :

A representation $\theta^* = \frac{H^*}{1-K^*}$ of $S^{-1}$ is a stable equilibrium of (13) if the A.R. part $1-K^*(z^{-1})$ is a positive real transfer function, i.e. $\mathrm{Re}(1-K^*(e^{i\omega})) > 0$ for $-\pi < \omega \leq \pi$.

We were not able to give global analysis of (13), which is not very surprising in view of the great complexity encountered in proving theorem 1. But we have a result which lies between global and local convergence, namely an evaluation of the domain in which $\theta^*$ is an attractor, under the assumption of theorem 2.

Let us consider the following candidate for Lyapunov function :

(17) $U(\theta) = 2\|\theta-\theta^*\|^2$.

We have (appendix 2), along the integral curves $(\theta_s)$ of (15) :

(18) $-\frac{d}{ds}U(\theta_s) = \mathbb{E}([(1-K^*(z^{-1}))e_t] \ (c_t - \alpha \operatorname{sgn} c_t))$,

with $c_t = {}^t\theta_s \ \phi_t$ and $e_t = c_t - a_t$ is the input error.

On the other hand, let us consider the following scheme

$$a_t \boxed{S(z^{-1})\ ?} x_t \boxed{\frac{1}{1-K^*(z^{-1})}} y_t \boxed{\begin{matrix}H(z^{-1})\\ \text{adjustable}\end{matrix}} c_t,$$

where $1-K^*$ is fixed and is the A.R. part of $\theta^*$ ; let us use the following algorithm for adjusting H

$$\begin{aligned} H_{t+1} &= H_t - \tau X_t(c_t - \alpha \operatorname{sgn} c_t) \\ &= H_t - \tau[(1-K^*(z^{-1}))Y_t].(c_t - \alpha \operatorname{sgn} c_t) \end{aligned}$$

where ${}^tH = (h_{-n},\ldots,h_{+n})$, ${}^tX_t = (x_{t+n},\ldots,x_t,\ldots,x_{t-n})$ and similary for $Y_t$. We have then, when $c_t = {}^tHY_t$

(20) $$\mathbb{E}(X_t(c_t - \alpha \operatorname{sgn} c_t)) = \overline{K}\cdot\mathbb{E}(Y_t(c_t - \alpha \operatorname{sgn} c_t)) = \overline{K}\cdot(\frac{d}{dH}J(H))$$

where the matrix ${}^t\overline{K}+\overline{K}$ is positive definite because the system $1-K^*(z^{-1})$ is strictly positive real ([5], [52]) ; hence the integral curves $(H_s)$ of the differential equation

(21) $$\frac{d}{dt}H = -\overline{V}(H), \quad \overline{V}(H) = \mathbb{E}(X_t(c_t - \alpha \operatorname{sgn} c_t)), \quad c_t = {}^tH\,Y_t$$

associated to (20) are such that $J(H_s)$ decrease to $\min J$, what ensures that

(22) $H_s \longrightarrow H^*$.

Now, let

(23) $\overline{U}(H) = 2\|H-H^*\|^2$,

we have that

(24) $$-\frac{d}{ds}\overline{U}(H_s) = \langle H_s - H, \overline{V}_{H_s}\rangle = \mathbb{E}([(1-K^*(z^{-1}))e_t] \times (c_t - \alpha \operatorname{sgn} c_t))$$

where $e_t = c_t - a_t$. Because of (22), there is a domain $\overline{D}(H^*)$ in which $-\frac{d}{dt}\overline{U}(H_t)$ is $>0$. Comparing (18) and (24), we see that

(25) $$\frac{d}{ds}U(\theta_s) = \frac{d}{ds}\overline{U}(H_s)$$

when $\theta_s$ and $(1/1-K^*) \circ H_s$ give the same system.

Now, let $D(\theta^*)$ the greatest ball centered at $\theta^*$ such that the corresponding H are in $\overline{D}(H^*)$, then we have

(26) $$\frac{d}{ds}U(\theta_s) < 0 \quad \text{for } \theta_s \in D(\theta^*).$$

This is more precise than local convergence because $\overline{D}(\theta^*)$ is generally not small ; but this does not ensure global stability because (23) is not a Lyapunov function for (21) ; in fact, even in the true gradient case (i.e. $K^*=0$), $\overline{U}$ is not a Lyapunov function ; appropriate Lyapunov functions are analyzed in [2], and are much more complicated, involving spherical coordinates.

## §5 : NUMERICAL RESULTS.

In both examples the input $(a_t)$ is uniformly distributed on the finite set $\{\pm1, \pm3, \pm5, \pm7\}$. Note that this distribution is not subgaussian in our sense because conditions A-2 are not satisfied here. But one can consider that this distribution is a good approximation of the uniform one on the interval [-7,+7] (See [2], [3] fo further details). Note that $\alpha = 5.25$.

Example 1 :

We compare RSGA and ERSGA with the classical RML and Extended Least squares algorithms (in stochastic gradient version [6]) on a minimum phase system : the lost with respect to those "optimal" second order algorithms is seen to be small.

$$x_t - \underbrace{0.8}_{=k_1} x_{t-1} + 0.5 x_{t-2} = a_t + \underbrace{0.7}_{h_1} a_{t-1} + 0.4 a_{t-2}$$

*Fig. 1* : Evolution of the mean square error for Extended Least squares and ERSGA.

*Fig. 2* : Evolution of the mean scuare error for RML and RSGA.

*Fig. 3* : Convergence of the estimates of $k_1$ and $h_1$ for RML and RSGA.

Example 2 :

We have taken an equivalent base-band telephone channel (non-minimum phase), which is sampled with $\Delta t = \frac{1}{3600}$ s, achieving a binary output of 9600 bits/s.

*Fig. 4* : Impulse response of the channel.

*Fig. 5* : Evolution of the mean square error and the number of errored data for RSGA.

## APPENDIX 1

Proof of (14).

We shall assume that the distribution f of $(a_n)$ is sufficiently smooth for the differentiation under the integrate to be justified. We have then (here $\delta$ is the symbolic Dirac function), because $c_t^* = a_t$

(A-1) $$\frac{d^2}{d\theta^2} J(\theta^\times) = \mathbb{E}(\phi_t K^* {}^t\phi_t K^* (1 - 2\alpha\ \delta(a_t))).$$

Now we have to calculate terms of the form

(A-2) $$\mathbb{E}(y_{t+u}\ z_{t+v}\ \delta(a_t)),$$

with $(y_t)$ and $(z_t)$ the outputs of some linear filter with input $(a_t)$ ; let us denote those filters in a M.A. form with infinitely many coefficients :

(A-3) $$y_t = \sum_{i\in\mathbb{Z}} \beta_i\, a_{t-i}, \quad z_t = \sum_{j\in\mathbb{Z}} \gamma_j\, a_{t-j}.$$

We have then

(A-4) $$\begin{aligned} &\mathbb{E}(y_{t+u}\ z_{t+v}\ \delta(a_t)) \\ &= f(0)\ \mathbb{E}(\sum_{i\neq u} \beta_i\, a_{t+u-i} \times \sum_{j\neq v} \gamma_j\, a_{t+v-j}). \end{aligned}$$

Using the fact that $(a_t)$ is i.i.d., we see that

(A-5) $$\sum_{i\neq u} \beta_i\, a_{t+u-i} = y_{t+u} - \mathbb{E}(y_{t+u}/a_t),$$

where $\mathbb{E}(\cdot/\cdot)$ denotes best linear prediction. Finally, (A-1), (A-4) and (A-5) give (14).

## APPENDIX 2

APPENDIX 2

$$-\frac{d}{ds} U(\theta_s) = {}^t(\theta_s - \theta^*)\ V(\theta_s).$$

Let $c_t$ and $\phi_t$ such that $c_t = {}^t\theta_s\, \phi_t$, then we have

$${}^t\theta_s V(\theta_s) = \mathbb{E}(c_t(c_t - \alpha\ \mathrm{sgn}\, c_t)),$$

and

$$\begin{aligned} {}^t\theta^* \phi_t &= k_1^* c_{t-1} + \ldots + k_p^* c_{t-p} + h_{-n}^* x_{t+n} + \ldots + h_n^* x_{t-n} \\ &= a_t + K^*(z^{-1}) e_t, \end{aligned}$$

hence

$$c_t - {}^t\theta^* \phi_t = (1 - K^*(z^{-1})) e_t,$$

and finally (18).

## REFERENCES


[1] A. BENVENISTE, M. GOURSAT, G. RUGET : *"A robust adaptive procedure for solving a Non Gaussian identification problem".* 8th IFIP Conf. on Opt. Tech., Würzburg Sept. 77 ; Lect. Notes in Control and Information Sciences 6, Springer V.

[2] A. BENVENISTE, M. GOURSAT, G. RUGET : *"Robust identification of a non-minimum phase system : blind adjustment of a linear equalizer in data communication".* To appear in IEEE A.C.

[3] A.B., M. BONNET, M.G., C. MACCHI, G.R. : Rapport LABORIA to appear.

[4] A.B., M.G., G.R. : *"Analysis of stochastic approximation schemes with applications to data communication algorithms".* To appear in IEEE A.C.

[5] L. LJUNG : *"On positive real Transfer Functions...".* IEEE AC-22 n° 4 pp. 539-550, 1977.

[6] L. LJUNG : *"Analysis of Recursive Stochastic Algorithms".* Ibidem pp. 551-575.

[7] T. SODERSTROM, L. LJUNG, I. GUSTAVSSON : *"A theoretical Analysis of Recursive Identification Methods".* Automatica Vol. 14 pp. 231-244, 1978.

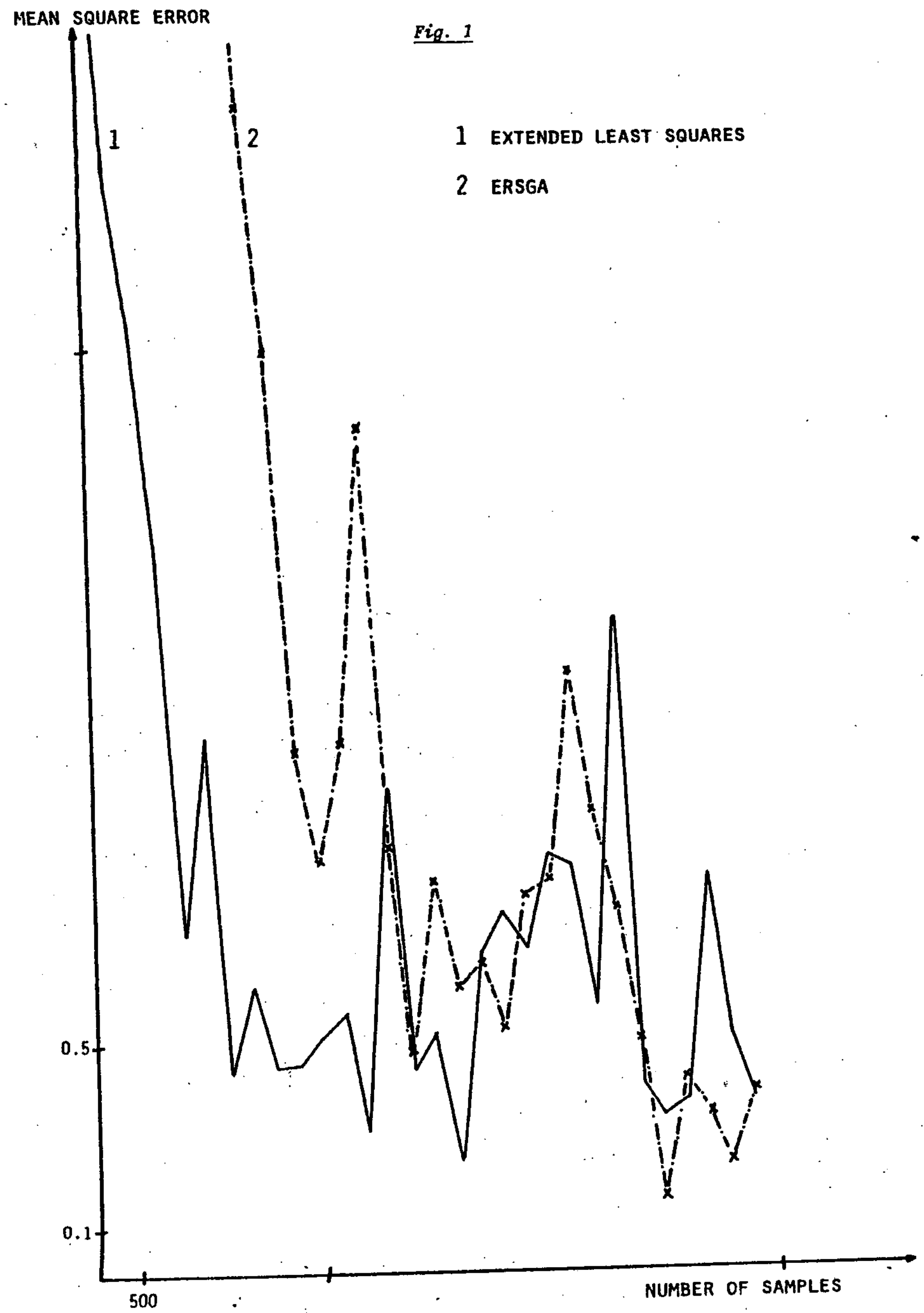

MEAN SQUARE ERROR
1 EXTENDED LEAST SQUARES
2 ERSGA
1
2
0.5
0.1
500
NUMBER OF SAMPLES

*Fig. 1*

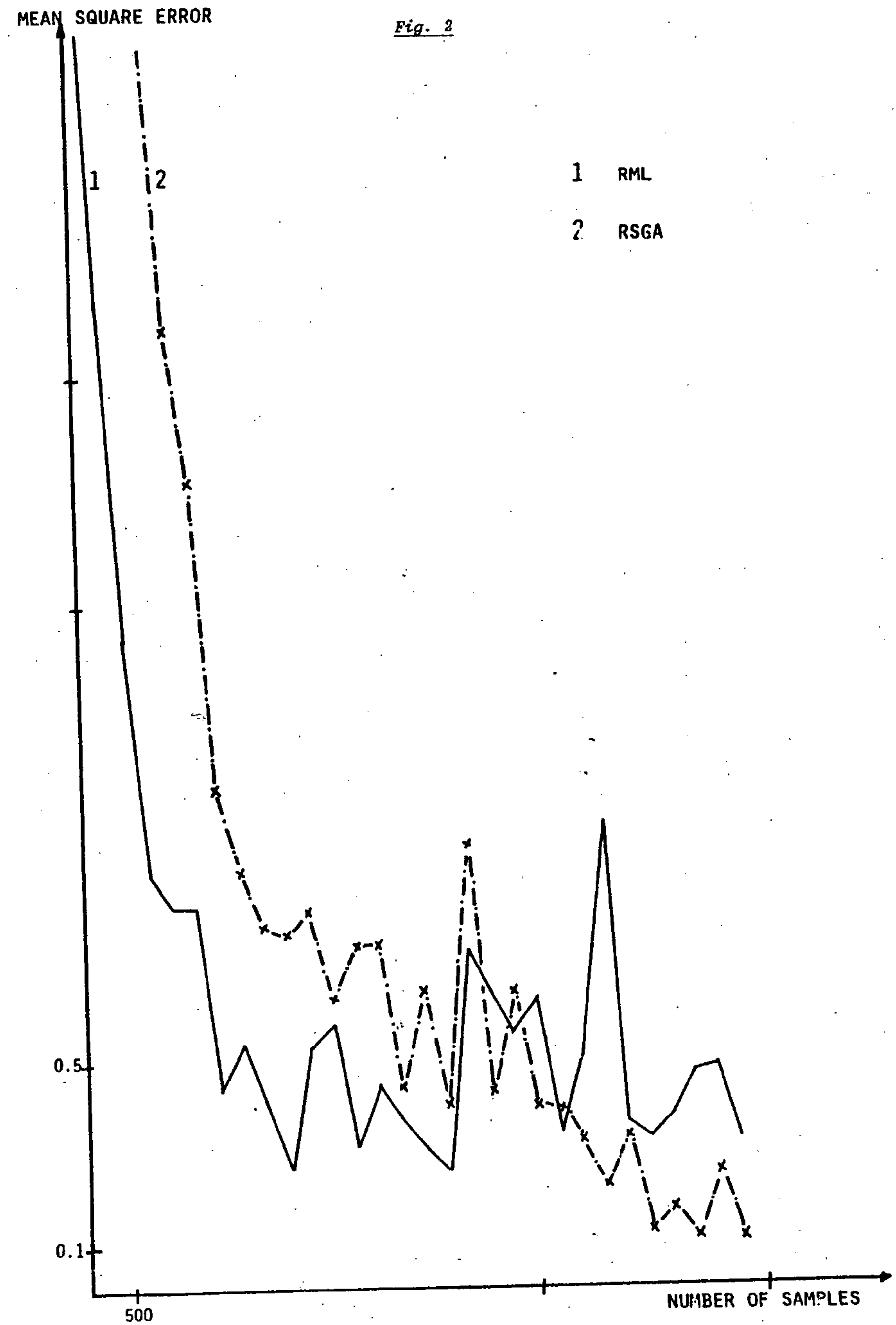

MEAN SQUARE ERROR
1 RML
2 RSGA
1
2
0.5
0.1
500
NUMBER OF SAMPLES

*Fig. 2*

*Fig. 3*

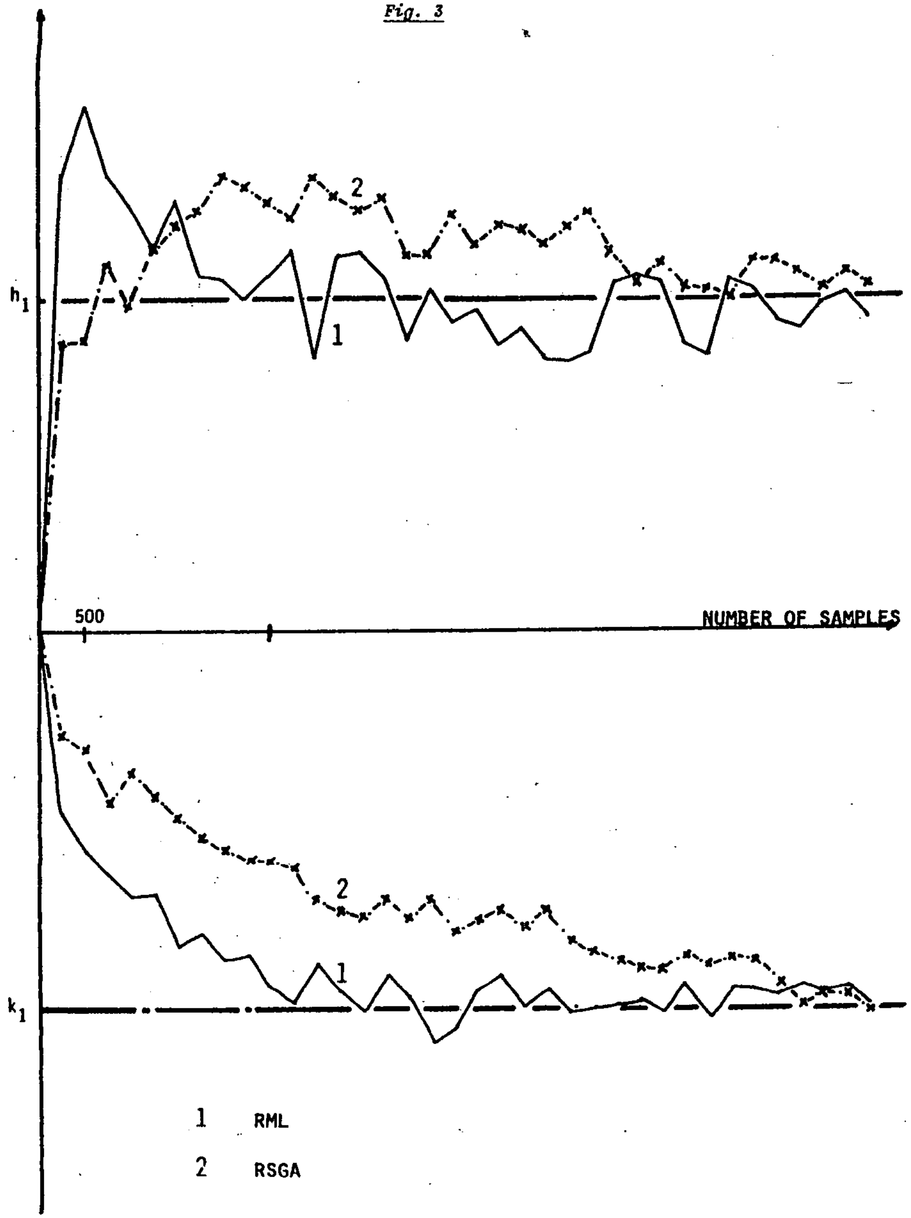



*Fig. 4*

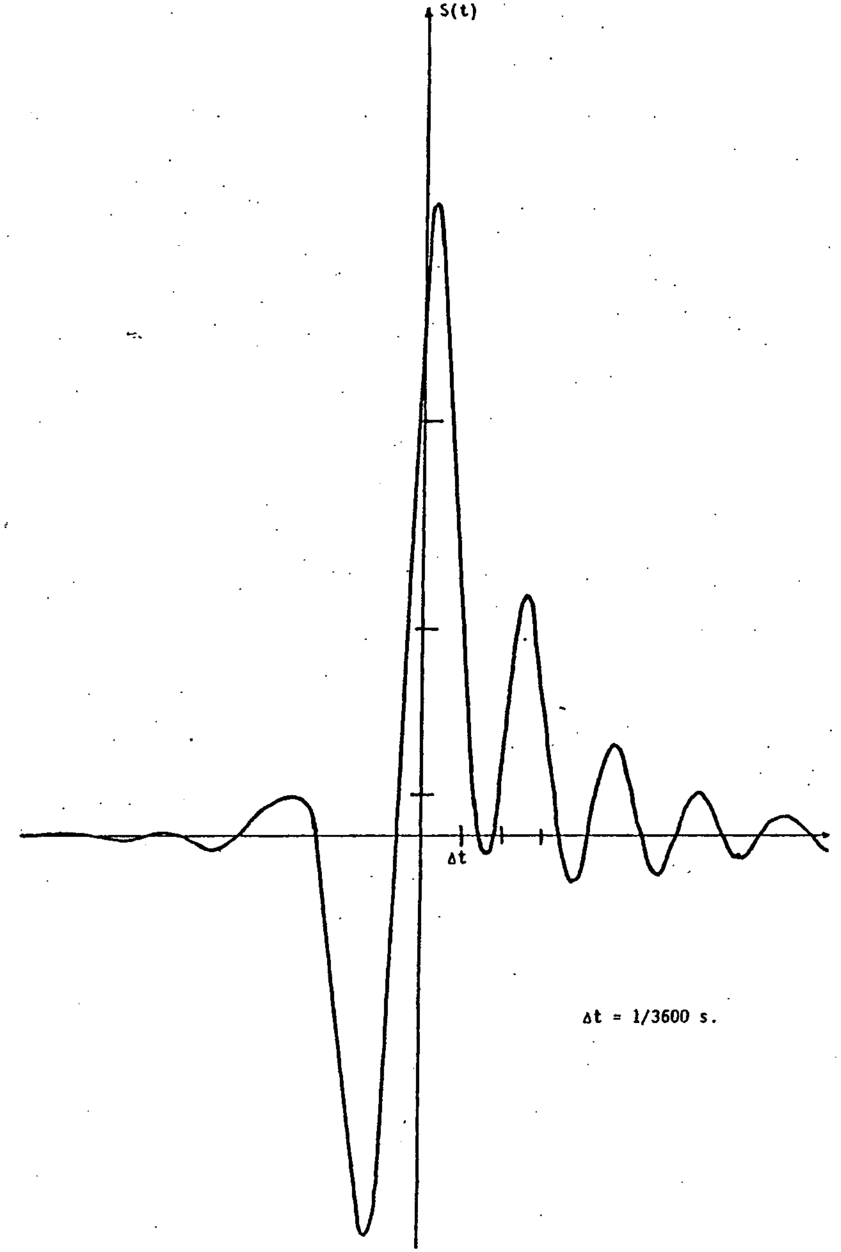
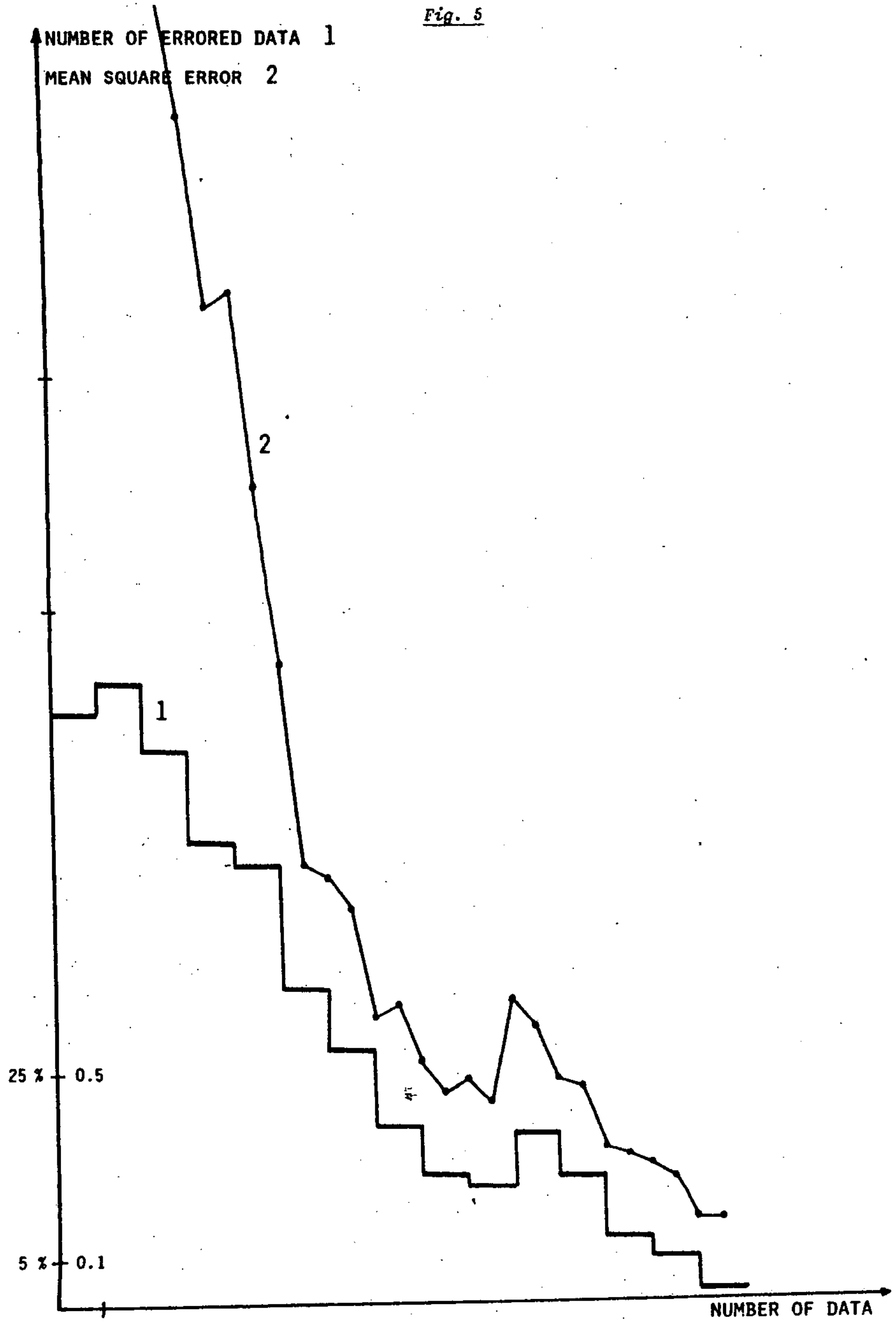


*Fig. 5*



NUMBER OF ERRORED DATA 1
MEAN SQUARE ERROR 2
2
1
25 % 0.5
5 % 0.1
NUMBER OF DATA